EXAMINING THE EFFECTS OF SITE-SELECTION CRITERIA FOR EVALUATING THE EFFECTIVENESS OF TRAFFIC SAFETY IMPROVEMENT COUNTERMEASURES

A Dissertation

by

PEI-FEN KUO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2012

Major Subject: Civil Engineering

Examining the Effects of Site-Selection Criteria for Evaluating the Effectiveness of Traffic Safety Improvement Countermeasures

EXAMINING THE EFFECTS OF SITE-SELECTION CRITERIA FOR EVALUATING THE EFFECTIVENESS OF TRAFFIC SAFETY IMPROVEMENT COUNTERMEASURES

A Dissertation

by

PEI-FEN KUO

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Dominique Lord
Committee Members,	Jeffrey Hart
	Luca Quadrifoglio
	Paul Nelson
Head of Department,	John Niedzwecki

May 2012

Major Subject: Civil Engineering

ABSTRACT

Examining the Effects of Site-Selection Criteria for Evaluating the Effectiveness of Traffic Safety Improvement Countermeasures. (May 2012)

Pei-Fen Kuo, B.E., National Taiwan University;
M.S., National Taiwan University
Chair of Advisory Committee: Dr. Dominique Lord

The before-after study is still the most popular method used by traffic engineers and transportation safety analysts for evaluating the effects of an intervention. However, this kind of study may be plagued by important methodological limitations, which could significantly alter the study outcome. They include the regression-to-the-mean (RTM) and site-selection effects. So far, most of the research on these biases has focused on the RTM. Hence, the primary objective of this study consists of presenting a method that can reduce the site-selection bias when an entry criterion is used in before-after studies for continuous (e.g. speed, reaction times, etc.) and count data (e.g. number of crashes, number of fatalities, etc.). The proposed method documented in this research provides a way to adjust the Naïve estimator by using the sample data and without relying on the data collected from the control group, since finding enough appropriate sites for the control group is much harder in traffic-safety analyses.

In this study, the proposed method, a.k.a. Adjusted method, was compared to commonly used methods in before-after studies. The study results showed that among all methods evaluated, the Naïve is the most significantly affected by the selection bias. Using the CG, the ANCOVA, or the EB method based on a control group (EB_{CG}) method can eliminate the site-selection bias, as long as the characteristics of the control group are

exactly the same as those for the treatment group. However, control group data that have same characteristics based on a truncated distribution or sample may not be available in practice. Moreover, site-selection bias generated by using a dissimilar control group might be even higher than with using the Naïve method. The Adjusted method can partially eliminate site-selection bias even when biased estimators of the mean, variance, and correlation coefficient of a truncated normal distribution are used or are not known with certainty. In addition, three actual datasets were used to evaluate the accuracy of the Adjusted method for estimating site-selection biases for various types of data that have different mean and sample-size values.

To my families

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Lord, and my committee members, Dr. Hart, Dr. Quadrifoglio, and Dr. Nelson, for their guidance throughout the completion of this dissertation.

I also want to extend my gratitude to my colleagues in Texas Transportation Institute (TTI). Special thanks are given to Dr. Troy, Dr. Mounce, Dr. Hawkins, and Dr. Eisele for their support for my research at TTI.

Finally, thanks to my families for their encouragement and love.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

Developing precise and reliable methods to evaluate countermeasure effectiveness is crucial, since erroneously measuring the safety effects could have drastic consequences both in terms of lives saved and funds wasted. This dissertation focuses on how site-selection criteria affect the accuracy of current methods of evaluating countermeasures for traffic safety improvement. Site-selection criteria can exercise a strong influence over the evaluation of these countermeasures; however, this influence has not been adequately addressed by the traffic safety community.

As described above, the primary purpose of this research is to describe how site selection effects can influence the evaluation of treatments and then present a new method that can remove or reduce site-selection effects when an entry criterion is used in before-after studies without relying on the use of a control group. This section consists of three sections. Section 1.1 provides the problem statement. In Section 1.2, specific objectives of this research are provided. The outline of the dissertation is presented in Section 1.3.

1.1 Problem Statement

Evaluating the effects of an intervention or a countermeasure on the number and severity of crashes or on the change of other surrogate measurements (e.g. driving speed, reaction time, and headway) is a very salient topic in highway safety. In fact, this topic has been researched thoroughly over the last 30 years (Abbess, 1981; Danielsson, 1986; Davis, 2000; Hauer, 1980a; Hauer, 1980b; Hauer et al., 1983; Hauer, 1997; Maher and Mountain, 2009; Miranda-Moreno, 2006; Miranda-Moreno, 2009; Wright et al., 1988). During this time, researchers have developed and applied various methods to minimize known biases associated with count and continuous data.

This dissertation follows the style of Accident Analysis and Prevention.

Over the years, we have seen a variety of methods that have been proposed for evaluating safety interventions. They include the Naïve before-after study, the before-after study with a control group, the Analysis of Covariance, the before-after study using the Empirical Bayes (EB) method, and more recently the before-after study using the full Bayes approach (Hauer, 1997; Hauer, 1984; Li et al., 2008; Park et al., 2010; Park and Lord, 2010; Persaud and Lyon, 2007). As an alternative to the before-after study, some people have suggested using a cross-sectional study (usually via a regression model) (Noland, 2003; Tarko et al., 1998). However, the before-after study is still considered the most appropriate methodology by most researchers, since it can directly account for changes that occurred at the sites investigated (Hauer, 1997).

One of the most important biases that have been documented in the literature which negatively influences the evaluation of treatments is the regression-to-the-mean (RTM). The RTM dictates that when observations characterized by very high (or low) values in a given time period and for a specific site (or several sites), it is anticipated that observations occurring in a subsequent time period are more likely to regress towards the long-term mean of a site (Hauer et al., 1983). Not including it could over-estimate the effects of the treatment (see, e.g., (Persaud, 2001). Although much work has been devoted to the RTM, very few studies have examined the selection bias on the effects of a treatment, at least as far as analyzing it as a distinct bias (see, e.g., (Davis, 2000; Hauer, 1980a). As discussed by (Cook and Wei, 2002; Davis, 2000)) and more recently by (Park and Lord, 2010), the site selection effects and RTM are distinct biases and influence the overall effectiveness of a treatment differently.

1.2 Research Objectives

The primary objective of this research was to describe how site selection effects can influence the evaluation of treatments. More specifically, the goal was to quantify the bias for the safety effectiveness of a treatment as a function of different entry criteria and other factors associated with traffic safety data. The study objective was accomplished using simulated data (supported by theoretical derivations documented in Appendix A and B) and observed data. In doing so, the following objectives were addressed in this study:

1. Examine how setting entry criteria in a trial affects the performance of particular traffic-safety countermeasures. This analysis used two types of safety datasets: discrete data (i.e., crash numbers, fatalities, and injuries) and continuous data (i.e., driving speed and driver-reaction time). Negative Binomial, normal distributions, and other common models were employed for this part of the analysis.
2. Estimate the difference in bias which is caused by using different entry-criteria values, and different methods of estimating countermeasure effectiveness. For count data, the most common before-after methods are: Naïve, using a control group (CG), EB method estimated using the method of moment (EB_{MM}), and EB method estimated using a control group (EB_{CG}). For continuous data, the most common methods are: the Naïve, CG, and the Analysis of Covariance (ANCOVA) methods. All possible influence factors (e.g. entry criteria, sample size, inverse-dispersion parameters, safety-effectiveness values, standard deviation of safety effectiveness, between-subject variances, and within-subject variances) will also be discussed. Bias on the estimates of the mean and variance of safety effectiveness will both be examined.
3. Derive the appropriate equations for estimating the performance of countermeasures (mean and variance) when the study includes entry criteria; and develop a suitable approach for adjusting other researchers' estimated values of treatment performance.

1.3 Outline of the Dissertation

This dissertation is divided into seven sections. Section 2 describes the background information about the RTM, site-selection effects, the truncated count model, hot-spot identification, and methods of estimating the effects of countermeasures. Section 3covers the approach used for conducting the simulation and calculation. Section 4 describes the results of the simulation analysis. Section 5 and 6 show the results for count and continuous data based on the observation data. The final Section, Section 7, summarizes the key study results and provides avenues for further research.

2. BACKGROUND

This section provides relevant background related to roadway- and highway-safety studies in five topics: 1) Regression-to-the-Mean (RTM); 2) Site-Selection Effects; 3) Truncated Models; 4) Hot-Spot Identification; and 5) Methods of Estimating Countermeasure Effectiveness.

2.1 Regression-to-the-Mean (RTM)

RTM refers to the concept that observations characterized by very high (or low) values in a given time period (K_1), and for a specific site, are anticipated to regress towards the long-term mean ($\bar{K}$) of a site in a subsequent time period (K_2) (Hauer, 1997). The characteristics of the RTM are illustrated in Figure 2-1.

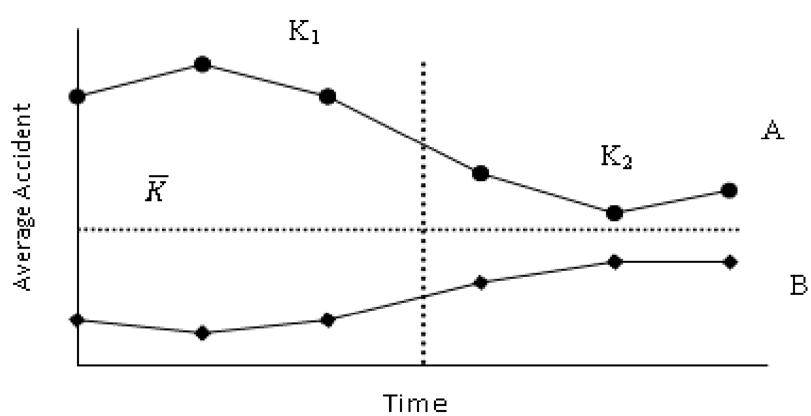


Figure 2-1 The RTM phenomenon in before-after study

The RTM is not new and was first observed more than a century ago by Francis Galton (Stigler, 1997). The RTM can be conceptualized mathematically using random variables in two time periods, labeled as 1 and 2, respectively. Let us assume that Y_1 and Y_2 are two random variables with almost exactly the same distribution, but where the conditional expectation $E[Y_2 | Y_1]$ is not equal to Y_1. It can be shown that the conditional expectation can be defined as a jointly normal distribution:

$$E[Y_2 | Y_1] = \rho Y_1 + (1-\rho)\mu \quad \textbf{(2.1)}$$

where ρ is the correlation between Y_1 and Y_2, and where μ is the common mean. Equation (2.1) shows that the RTM effect is a function of the correlation coefficient. When the correlation coefficient is equal to 1, there is no RTM, since $E[Y_2 | Y_1] = Y_1$. On the other hand, when the correlation coefficient is not equal to 1, we observe the presence of the RTM. Smaller values of ρ are associated with larger RTM effects. Equation (2.1) also shows that the magnitude of the RTM can be computed by taking the difference of $E[Y_2 | Y_1]$ and Y_1 (Figure 2-2).

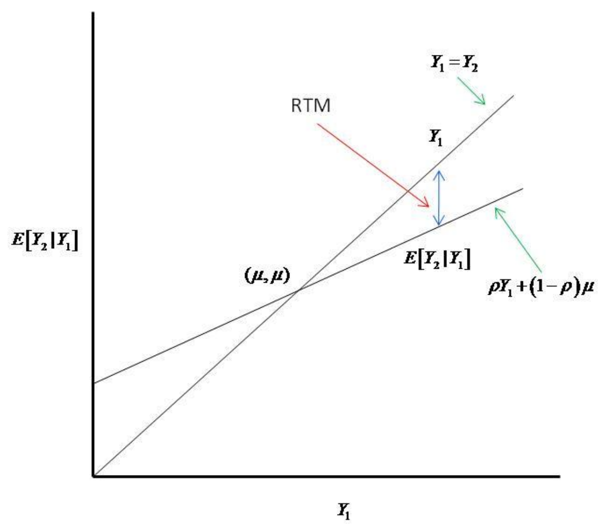


Figure 2-2 Graphical representation of the RTM phenomenon

Hence, estimating countermeasure effectiveness by a simple before-after comparison–without considering the RTM effects can be risky. Numerous studies have already discussed this problem (Hauer, 1997; Pendleton, 1991; Persaud, 2001). The studies claim that the EB approach can remove RTM bias by providing a more accurate estimator of the expected accident count. According to the theory, the RTM is explained by the temporal correlation (ρ) for observations that are evaluated at different time periods (Chuang-Stein and Tong, 1997; Stigler, 1997). RTM will exist unless there is a perfect correlation, $\rho = 1$. The smaller the correlation, the greater the RTM.

Although previously published documents have often confused RTM with selection bias, the researcher suggests that RTM should be separated from site-selection effects, because RTM bias (Hauer, 1997) may exist even when entities are not selected by their unusually high or low response (Cook and Wei, 2002; Hauer, 1997). In others words, the RTM is defined as a natural phenomenon caused by its extreme value in its first measurement; however, site-selection effects are artificial impacts caused by how entry criteria are set. In this study, the goal is to examine how entry-criteria values affect the evaluation result. This examination may then serve as a reference for traffic engineers deciding what entry criteria to use during the experiments.

2.2 Site-Selection Effects

This section is divided into four subsections. Section 2.2.1 briefly outlines the characteristics of site-selection bias. Section 2.2.2 and section 2.2.3 describe the site-selection bias for continuous data and count data. Section 2.2.4 summaries main findings and problems of site-selection site from current studies.

2.2.1 General Characteristics

As discussed in Park and Lord (2010) and in the references herein, site-selection biases and RTM are two different biases. The general idea of site-selection effects is to set the entry criteria so as to convert the original population distribution to a truncated sample distribution, which results in a change in the unbiased estimators of mean and variance.

Ignoring these changes will cause a bias in estimating the countermeasure effectiveness (Figure 2-3).

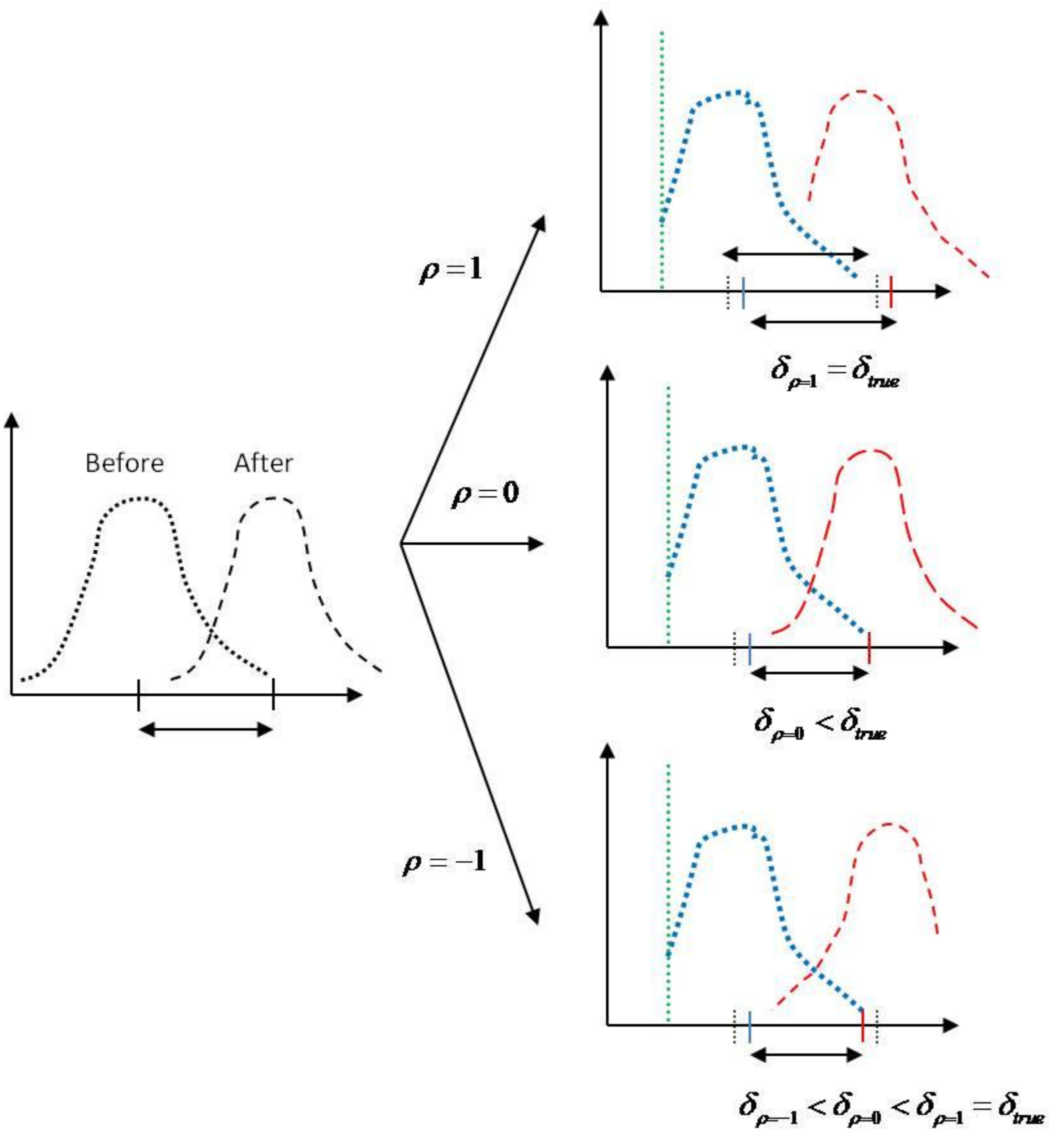


Figure 2-3 The distribution of population and the truncated sample

In Figure 2-3, the left-hand side shows the probability related to the normally distributed data in the *before* and *after* periods (without selection). The difference between the mean of these two curves is defined as δ_{true}. After setting the minimum entry criteria (the dashed vertical line), the data distribution in the *before* period is left-truncated, as indicated by the blue curved line on the right-hand side. It should be noted that setting the same entry criteria might cause different effects according to different correlation coefficient (ρ) values with the *before* and *after* data. If ρ is equal to 1, the estimator of the difference, δ_{true}, is unbiased because the mean value increases in the same manner in the *before* and *after* periods. If ρ is equal to 0, the naïve estimator of difference is less than its true value. The mean in the *before* period increases, but the mean in the *after* period remains the same because the before-after data are independent. If ρ is negative, the estimator of the difference might become lower than the above values. Removing the data with low values in the *before* period may also remove data with high values in the *after* period because of the negative correlation. Hence, the difference becomes much lower because of the higher mean in the *before* period and a lower mean in the *after* period.

Table 2-1 summarizes the mathematical equations used for quantifying the RTM and site-selection effects (i.e. data are left-truncated). It should be noted that the probability of the data being characterized by the site-selection effects in the *after* period is equal to the truncated normal distribution multiplied by the conditional normal distribution of the *after* period (Cook and Wei, 2002).

Table 2-1 Equations Describing Site-Selection and RTM Effects.

Effects	Before	After
Site selection	$P(Y_1 \| Y_{i1} > C), i : site\ i$	$P(Y_2 \| Y_{i1} > C)$ $= P(Y_2 \| Y_{i1}) P(Y_{i1} > C)$
Regression-to-the-Mean	$P(Y_1 \| Y_{i1}), Y_{i1} > \mu_1$	$P(Y_2 \| Y_{i1}), Y_{i1} > \mu_1$

While discussion surrounding RTM has been extensive, estimating the magnitude of site-selection effects according to their entry-criteria values is a new topic in traffic safety. It is not, however, a new topic within the realm of scientific inquiry. Medical studies – particularly clinical trials – frequently employ a site-selection-based approach to analysis. Usually, the entry criteria are chosen based on historical or baseline measurements of clinical signs and symptoms, which serve as the response variables. Cook and Wei (2002), for instance, discussed the possible impacts of selection effects on testing new medicines in clinical studies. They derived equations for estimating the bias linked to the treatment performance when the response variables followed a normal distribution and were classified as discrete data. The treatment performances here are categorized into two types: 1) The difference between the *before* and *after* periods of an average response ($\delta= \mu_2-\mu_1$); and, 2) The ratio of the *before* and *after* periods of an average response ($\theta= \mu_2/\mu_1$). Two data sets, from an epilepsy trial and from a myocardial ischemia, were used to illustrate the effects of ignoring the selection mechanism when there were high, medium or low entry-criteria levels. Their most important findings included the following:

1) For Normal Distribution responses: When the scientists set the entry criteria for choosing experiment subjects, the original unbiased estimators for treatment effectiveness become biased. Even when there is no relationship between the response in the *before* and *after* periods the analysis can be biased. Furthermore, when the treatment does not work ($\mu 1=\mu 2$), using the naïve method may result in a positive estimation of treatment effectiveness, especially for low-value responses. The selection bias will exist until the correlation (ρ) is 1 or the entry criteria (C) are close to $-\infty$. However, when a control group is used, this bias will no longer exist.

2) For Count Data responses: The result is analogous to the Normal Distribution response. The bias for estimating the treatment performance increases when entry

criteria are set higher; however, the bias will continue to exist until the entry criteria are less than 0.

In sum, Cook and Wei showed that setting higher entry criteria may cause a larger bias in estimating the performance of the treatment; they also showed that using a naïve before-after method tends to cause an overestimation of treatment performance when the response number is low. The overall indication is that site-selection effects can play an integral role in improving traffic safety, as the distributions of crash frequency usually have a low mean (Lord and Mannering, 2010). However, Cook and Wei's theory is not without its flaws. There are three major problems that need to be resolved before we apply Cook and Wei's study results to the realm of traffic safety. First, there are several typos in bias-estimation equations of effectiveness and variance. Secondly, the calculation of the Control Group method for count data in medical studies is different from when it is used in traffic safety studies. Furthermore, Cook and Wei did not consider the *dissimilar* control group – a very common control group type in real situations – and examine the accuracy of their bias equation when the parameters (e.g., α, Λ_1, ρ) are unknown. To these ends, the following sections show the updated equation of bias caused by the site-selection criteria.

2.2.2 Site Selection Effects for Normal Distribution Responses

The equations shown below are adapted from Cook and Wei (2002). These equations may be used for analyzing travel speed, driver-reaction time, and other data that follow a normal distribution. Let Y_{i1} and Y_{i2} denote the response of subject i in the *before* (k=1) and the *after* (k=2) periods. Let the sample size be n. The initial assumption is that the response of subject i can be separated by three components: time (μ_k, k=1 or 2), subject (u_i, i=1,..,n), and random effects (e_{ik}, k=1 or 2, i=1,..,n).

$$Y_{ik} = \Lambda_k + u_i + e_{ik} \tag{2.2}$$

where $u_i \sim N(0, \phi)$ and $e_{ik} \sim N(0, \sigma^2)$.

In addition, u_i and e_{ik} are considered independent. Let $Y_{ik} = (Y_{i1}, Y_{i2})'$ be a bi-variables normal distribution with $E(Y_{ik}) = \Lambda_k$, $Var(Y_{ik}) = \sigma^2 + \phi$, and with the correlation of Y_{i1}, Y_{i2} $(\rho_{Y_{i1},Y_{i2}}) = \frac{\phi}{(\sigma^2+\phi)}$. To see the difference in response between the *before* and *after* periods, estimators with and without entry criteria are shown in sections 2.2.2.1 and 2.2.2.2.

2.2.2.1 The Estimator of μ_k and δ, γ Without Entry Criteria

According to the method of moment, the unbiased estimator of the mean response is:

$$\hat{\mu}_k = \overline{Y_k} = \frac{\sum_{i=1}^{n} Y_{ik}}{n} \tag{2.3}$$

Hence, the effectiveness of the countermeasure could be the difference between the responses in the *before* and *after* periods (δ), or the ratio of the responses in the *before* and *after* period (θ):

$$\hat{\delta} = \hat{\Lambda}_2 - \hat{\Lambda}_1 = \overline{Y}_2 - \overline{Y}_1 \tag{2.4}$$

$$\hat{\theta} = \frac{\hat{\Lambda}_2}{\hat{\Lambda}_1} = \frac{\overline{Y}_2}{\overline{Y}_1} \tag{2.5}$$

Usually, comparison analyses use a paired t-test for examining whether δ is different from zero, or whether θ is different from 1. If δ is smaller than 0, or if θ is less than 1, then traffic engineers declare that this treatment reduces the response rate.

2.2.2.2 The Estimator of μ_k and δ, γ With Entry Criteria

With entry criteria, the distribution of Y_{ik} is truncated by *C*. Hence, the original estimator of the mean response in equation (2.3) is no longer unbiased, and the unbiased estimator of the mean response in the *before* period becomes:

$$E\left(Y_{i1} \middle| Y_{i1} > C\right) = \mu_1 + \sqrt{\sigma^2 + \phi} \times \frac{f(d)}{(1 - F(d))} \tag{2.6}$$

where $d = \frac{(c - \Lambda_1)}{\sqrt{\sigma^2 + \phi}}$ and f(d) and F(d) are the PDF and CDF of standard normal distribution respectively. Using the Bayesian theorem, the estimator for the response in the *after* period can be calculated as follow:

$$E\left(Y_{i2} | Y_{i1} > C\right) = \Lambda_2 + \rho\left(Y_{i1} - \Lambda_1\right) \tag{2.7}$$

And the estimator for δ is:

$$\hat{\delta} = E\left(Y_{i2} - Y_{i1} | Y_{i1} > C\right) = \Lambda_2 - \Lambda_1 + \left(\rho - 1\right)\left(E(Y_{i1} | Y_{i1} \geq C\right) - \Lambda_1)$$

$$= \Lambda_2 - \Lambda_1 + (\rho - 1) \times \sqrt{\sigma^2 + \phi} \frac{f(d)}{(1 - F(d))} \tag{2.8}$$

Cook and Wei used equation (2.8) to show that even if the responses in the *before* and *after* periods are independent (ρ =0), the bias still exists. However, if the entry criteria are close to negative infinity then f (d) tends towards zero or ρ =1, and the bias may be ignored. Also, $\hat{\delta}$ exists even the treatment does not work ($\Lambda_2 = \Lambda_1$).

The explanations from Cook and Wei are correct but not clear enough, so this study has added one estimator equation (2.9) for θ. There are two reasons for doing so: (1) traffic engineers usually prefer using θ to represent treatment performance (re: CMF in the *Highway Safety Manual*); and, (2) there can be different boundary conditions for θ and ρ.

$$\hat{\theta} = E\left((Y_{i2} / Y_{i1} | Y_{i1} > C\right) = \frac{\Lambda_2 + \rho \times \sqrt{\sigma^2 + \phi} \times f(d) \big/ (1 - F(d))}{\Lambda_1 + \sqrt{\sigma^2 + \phi} \times f(d) \big/ (1 - F(d))} \tag{2.9}$$

In equation (2.9), the bias will not exist when $f(d) \big/ (1 - F(d))$ is very small, or when ρ is Λ_2 / Λ_1. If ρ decreases further, or if the distance between the entry criteria and the mean is larger, then the bias also becomes larger. It should be noted that ρ =1 can make $\hat{\delta}$ unbiased but $\hat{\theta}$ still biased until $\Lambda_2 = \Lambda_1$.

2.2.2.3 The Estimator of μ_k and δ, γ with Entry Criteria with Control Group

One subscript is added to index the treatment group (j=1) and the control group (j=2). In medical studies, the estimator for δ is unbiased when $\Lambda_{11} = \Lambda_{21} = \Lambda_{22}$, $\rho_1 = \rho_2 = \rho$ and $d_1 = d_2 = d$

$$\begin{aligned}\hat{\delta}_{CG} &= E\left((Y_{i12} - Y_{i11} | Y_{i11} > C\right) - E\left((Y_{i22} - Y_{i21} | Y_{i21} > C\right) \\ &= \Lambda_{12} - \Lambda_{11} + \left(\rho_1 - 1\right)\left(E(Y_{i11} | Y_{i11} > C\right) - \Lambda_{11}) - \left[\Lambda_{22} - \Lambda_{21} + \left(\rho_2 - 1\right)\left(E(Y_{i21} | Y_{i21} > C\right) - \Lambda_{21})\right]\end{aligned} \tag{2.10}$$

where $\left(E(Y_{i11} | Y_{i11} > C\right) - \Lambda_{11}) = \sqrt{\sigma^2 + \phi} \times {f(d_1)} \big/ {(1 - F(d_1))}$

However, the estimator for θ is biased even when $\Lambda_{11} = \Lambda_{21} = \Lambda_{22}$, $\rho_1 = \rho_2 = \rho$ and $d_1 = d_2 = d$. The selection bias still exists until ${f(d)} \big/ {(1 - F(d))}$ is very small or $\rho = 0$.

$$\begin{aligned}\hat{\theta}_{CG} &= \frac{E\left((Y_{1i2} / Y_{i11} | Y_{i11} \geq C\right)}{E\left((Y_{i22} / Y_{i21} | Y_{i21} \geq C\right)} = \frac{\dfrac{\Lambda_{12} + \rho_1 \times \sqrt{\sigma^2 + \phi} \times {f(d_1)} \big/ {(1 - F(d_1))}}{\Lambda_{11} + \sqrt{\sigma^2 + \phi} \times {f(d_1)} \big/ {(1 - F(d_1))}}}{\dfrac{\Lambda_{22} + \rho_2 \times \sqrt{\sigma^2 + \phi} \times {f(d_2)} \big/ {(1 - F(d_2))}}{\Lambda_{21} + \sqrt{\sigma^2 + \phi} \times {f(d_2)} \big/ {(1 - F(d_2))}}} \\ &= \frac{\Lambda_{12} + \rho_1 \times \sqrt{\sigma^2 + \phi} \times {f(d)} \big/ {(1 - F(d))}}{\Lambda_{11} + \rho_2 \times \sqrt{\sigma^2 + \phi} \times {f(d)} \big/ {(1 - F(d))}}\end{aligned} \tag{2.11}$$

Based on equations (2.10) and (2.11), it is very clear that using the CG method with the exact same control-group data may remove site-selection bias for $\hat{\delta}$. However, $\hat{\theta}_{CG}$ is still biased until entry criteria is close to -∞, $f(d) / (1-F(d))$ is very small, or if the responses in the *before* and *after* periods are independent ($\rho = 0$).

2.2.3 Site Selection Effects for Count Data

The theorem of how site selection affects count data is analogous to that for continuous data. Suppose that the site i (i=1, 2… m) has N_{i1} crashes in the before period (time length = t1) and the crash counts of site i in the after period (time length = t2) is N_{i2}. Let N_{ik} follow the Poisson distribution ($N_{ik} \sim$ Poisson ($u_{it}\Lambda_k$)), where u_{it} is the subject-specific random effect, and where Λ_k is the average crash rate ($\Lambda_k = \lambda_{it} \times t_k$). λ_{it} is the instant rate of crash.

2.2.3.1 The Estimators of μ_k and δ, γ Without Entry Criteria

Suppose that there are no entry criteria for choosing experimental subjects, and that the expected value of the mean is:

$$E(N_{ik}) = u_i \Lambda_k \tag{2.12}$$

If we assume $u \sim \mathrm{N}(1, \alpha)$, then the estimator of response, variance and covariance are:

$$E(N_{ijk}) = \Lambda_k \tag{2.13}$$

$$Var\left(N_{ik}\right) = \Lambda_k + \Lambda_k^{\ 2} \tag{2.14}$$

$$Cov\left(N_{i1}, N_{i2}\right) = \Lambda_1 \Lambda_2 \tag{2.15}$$

Then, the marginal distribution, $P(N_{ik}, ?_{k}\ \alpha)$, is the Poisson-Normal model.

If we assume $u \sim$ Gamma (α^{-1}, α), then the estimator of response and variance, and the probability function become:

$$E(N_{ik}) = u_i \Lambda_k = \frac{\alpha^{-1}(\frac{\Lambda_k \alpha}{1+\Lambda_k \alpha})}{(\frac{1}{1+\Lambda_k \alpha})} = \Lambda_k \tag{2.16}$$

$$Var(N_{ik}) = \Lambda_k(1+\Lambda_k \alpha) \tag{2.17}$$

From the above equation, we can see that (2.16) and (2.17) are consistent with (2.13) and (2.14). Then, $P(N_{ik}, \Lambda_k, \alpha ?)$ is the Poisson-gamma model (Negative- Binomial model) and its join distribution is:

$$P(N_{i1}, N_{i2}, \Lambda_1, \Lambda_2, \alpha) = \frac{\Gamma(\alpha^{-1} + N_{i1} + N_{i2})}{\Gamma(\alpha^{-1}) N_{i1}! N_{i2}!} (\frac{1}{1+(\Lambda_1+\Lambda_2)\alpha})^{\alpha^{-1}+N_{i1}+N_{i2}} (\Lambda_1 \alpha)^{N_{i1}} (\Lambda_2 \alpha)^{N_{i2}}, k = 1, 2 \tag{2.18}$$

When $\theta = (\Lambda_1, \Lambda_2, \alpha)$, we can derive the log-likelihood function from (2.18)

$$l(\theta) = \sum_{i-1}^{m} \left[\sum_{l=1}^{Ni.} \log(\alpha^{-1} + l - 1) + \sum_{k=1}^{2} n_{ik} \log(\Lambda_k) + N_{i.} \log \alpha - (\alpha^{-1} + Y_{i.}) \log(1 + \Lambda . \alpha) \right] \tag{2.19}$$

Then, we can obtain the score vectors by partial differential of the above log-likelihood function for obtaining the MLE estimates.

$$S_1(\theta) = \frac{N_{.1}}{\Lambda_1} - \frac{m + N_{..}\alpha}{1+\Lambda_{.}\alpha} \tag{2.20}$$

$$S_1(\theta) = \frac{N_{.2}}{\Lambda_2} - \frac{m + n_{..}\alpha}{1+\Lambda_{.}\alpha} \tag{2.21}$$

$$S_3(\theta) = \sum_{i=1}^{m}\left[\sum_{l=1}^{Y_{i.}} \frac{1}{\alpha+(l-1)\alpha^2} + \frac{Ni.}{\alpha} + \frac{\log(1+\Lambda_{.}\alpha)}{\alpha^2} - \frac{(1+N_{i.}\alpha)\Lambda_{.}}{\alpha+\Lambda_{.}\alpha^2}\right] \tag{2.22}$$

Let S(θ) = 0. The estimator for maximum likelihood of response, and the difference of response between the *before* and *after* periods are:

$$\hat{\Lambda} = N_{\cdot k} / m, k = 1, 2 \tag{2.23}$$

$$\hat{\delta} = \hat{\Lambda}_2 - \hat{\Lambda}_1 = \overline{N}_2 - \overline{N}_1 \tag{2.24}$$

$$\hat{\theta} = \frac{\hat{\Lambda}_2}{\hat{\Lambda}_1} = \frac{\overline{N}_2}{\overline{N}_1} \tag{2.25}$$

Based on equations (2.23)-(2.25), the estimators for δ and θ are still unbiased. Their equations are also the same as equations (2.3) and (2.4).

2.2.3.2 The Estimators of μ_k and δ, γ With Entry Criteria

Suppose that site i has *C* or more crashes in the *before* period (time length = t_1, $N_{i1}>C$). Equation (2.18) then does not work when the trial contains entry criteria because the sample model is truncated. The new distribution is:

$$P(N_{i2} \mid N_{i1}) = \frac{\Gamma(\alpha^{-1}+N_{i.})}{\Gamma(\alpha^{-1}+N_{i1})N_{i2}!} \frac{(1+\Lambda_1\alpha)^{\alpha^{-1}+N_{i1}}(\Lambda_2\alpha)^{N_{i2}}}{(1+\Lambda_1\alpha)^{\alpha^{-1}+N_{i1}}} \tag{2.26}$$

Using the moment generation function, the score function is:

$$E\left\{S_k(\theta) \middle| N_{i1} \geq C, i = 1, \ldots, m\right\} = \frac{m\mu_k}{\Lambda_k} - \frac{m(1+(\mu_1+\mu_2)\alpha)}{1+\Lambda.\alpha}, k = 1, 2 \tag{2.27}$$

where $\mu_k = E(N_{ik} \mid N_{i1} > C)$

Similar to what was discussed in section 2.2.3.1, we can set S(θ)=0 to get the maximum likelihood estimators of Λ_1 and Λ_2. Then, the biases of the estimators of Λ_1 and Λ_2 are :

$$\lim_{m\to\infty} E(\hat{\Lambda}_1 - \Lambda_1 \mid N_{i1} > C, i = 1,\cdots,m) = \mu_1 - \Lambda_1 \tag{2.28}$$

$$\lim_{m\to\infty} E(\hat{\Lambda}_2 - \Lambda_2 \mid N_{i1} > C, i = 1,\cdots,m) = \frac{\Lambda_2\alpha(\mu_1 - \Lambda_1)}{\Lambda_1\alpha + 1} \tag{2.29}$$

Moreover, we can use (2.28) minus (2.29) to get the bias of the estimator of δ,

$$\lim_{m\to\infty} E(\hat{\delta} - \delta \mid N_{i1} > C, i = 1,\cdots,m) = (\mu_1 - \Lambda_1)\left[\frac{\delta\alpha - 1}{\Lambda_1\alpha + 1}\right] \tag{2.30}$$

The same procedure was used to obtain the bias of the estimator of θ, as seen in equation (2.31). However, equations (2.30) and (2.31) were both modified to account for typos found in the original equation proposed by Cook and Wei. Please see the red color for the corrections. The first typo affects the magnitude of site-selection bias, and the second typo changes the bias from negative to positive.

$$\lim_{m\to\infty} E(\hat{\theta} - \theta \mid N_{i1} > c, i = 1,\cdots,m) = -\theta\left[\frac{(\mu_1 - \Lambda_1)}{\mu_1(\Lambda_1\alpha + 1)}\right] \tag{2.31}$$

This equation is very important for estimating the site selection effect, and we will discuss it further in section 2.2.4.

2.2.3.3 The Estimators of μ_k and δ, γ With Entry Criteria And With Control Group

Using a similar equation to that in section 2.2.3.2, we only need to add one extra subscript, j, to distinguish the comparison group from the control group. The term j is equal to 2 for the treatment group and 1 for the control group. The other conditions and assumptions are the same as those in Section 2.2.3.2. As such, the responses still follow the Poisson distribution ($Y_{ijk} \sim Poission(u_{ij}\Lambda_{jk})$), and u_{ij} follows the Gamma distribution (α^{-1}, α). $\Lambda_{jk} = \lambda_{jk} \times \tau_k$ is the average population count for the *k*th period. The estimator of the average crash rate ($\hat{\Lambda}_{jk}$) is $\sum_i N_{ijk} / n$. According to Hauer (1997), the equation that traffic engineers use to calculate the effectiveness is: $\delta_{CG} = \Lambda_{22} - \Lambda_{21}\frac{\Lambda_{12}}{\Lambda_{11}}$ instead of $\delta = (\Lambda_{22} - \Lambda_{21})\text{-}(\Lambda_{12} - \Lambda_{11})$ in Cook and Wei's paper. The MLE estimator of δ_{true} should be changed as shown below:

$$\delta_{CG} = \Lambda_{22}\left(\frac{\mu_{21}\alpha+1}{\Lambda_{21}\alpha+1}\right) - \mu_{21}\frac{\Lambda_{12}\left(\frac{\mu_{11}\alpha+1}{\Lambda_{11}\alpha+1}\right)}{\mu_{11}} \tag{2.32}$$

where $\mu_{jk} = E(N_{ijk} \mid N_{ij1} > C)$. If we put in the optimal assumption $\Lambda_{11} = \Lambda_{21} = \Lambda_1$ and $\mu_{11}=\mu_{21}=\mu_1$, then equation (2.32) is simplified to:

$$\hat{\delta}_{CG} = \frac{(\Lambda_{22} - \Lambda_{12})(\mu_1\alpha+1)}{(\Lambda_1\alpha+1)} \tag{2.33}$$

or equivalently to

$$\hat{\delta}_{CG} - \delta = \frac{(\Lambda_{22} - \Lambda_{12})(\mu_1\alpha+1)}{(\Lambda_1\alpha+1)} - (\Lambda_{22} - \Lambda_{12}) = \frac{\alpha(\Lambda_{22} - \Lambda_{12})(\mu_1 - \Lambda_1)}{(\Lambda_1\alpha+1)} \tag{2.34}$$

From equation (2.34), it should be clear that the MLE estimator of δ is still biased except in three conditions: when the dispersion parameter is zero; if the treatment does not work; or when it does not contain entry criteria. This bias will increase as entry criteria become larger. Following the same steps, the estimator of θ should become:

$$\hat{\theta}_{CG} - \theta = \frac{\Lambda_{22}(\frac{\mu_{21}\alpha+1}{\Lambda_{21}+1})}{\frac{\mu_{21}}{\mu_{11}} \times \Lambda_{12}(\frac{\mu_{11}\alpha+1}{\Lambda_{11}+1})} - \frac{\Lambda_{22}}{\Lambda_{21} \times \Lambda_{12}/\Lambda_{11}} \tag{2.35}$$

Let's use the optimal assumptions $\Lambda_{11} = \Lambda_{21} = \Lambda_1$ and $\mu_{11}=\mu_{21}=\mu_1$. Then, equation (2.35) is simplified as:

$$\hat{\theta}_{true} - \theta_{true} = \frac{\Lambda_{22}(\frac{\mu_1\alpha+1}{\Lambda_1+1}) \Big/ \mu_1}{\Lambda_{12}(\frac{\mu_1\alpha+1}{\Lambda_1+1}) \Big/ \mu_1} - \frac{\Lambda_{22}/\Lambda_1}{\Lambda_{12}/\Lambda_1} = 0 \quad (2.36)$$

Based on equation (2.36), it can be assumed that using a control group may cause the estimator of θ to become asymptotically unbiased when $\Lambda_{11} = \Lambda_{21} = \Lambda_1$ and $\mu_{11}=\mu_{21}=\mu_1$. However, this situation is very rare because two initial assumptions need to be present: optimal control group and large sample size. Moreover, more conditions are necessary to make the estimator of δ unbiased.

2.2.3.4 Site Selection Effects for Dispersion Parameter

The dispersion parameter may be another interesting parameter. Based on (Cook and Wei, 2002) study, site selection effect also biases its equation as shown below:

For estimating the bias of the dispersion parameter, we can use a quasi-likelihood approach. Equation (2.37) was corrected from Cook and Wei's result. As shown in equation (2.39), setting an entry criterion may lead to an underestimated dispersion parameter. Interestingly, Ye and Lord (2009) using a different approach, also discussed underestimating the variance in before-after studies.

$$\alpha^* = \mu_2^{-1}\left(\frac{v_2}{\mu_2}\frac{m-1}{m} - 1\right) \quad (2.37)$$

$$\alpha^* = \alpha\left[\frac{1+\alpha(v_1+\mu_1)}{(1+\mu_1\alpha)^2}\right] \quad (2.38)$$

$$\lim_{x\to\infty} E(\hat{\alpha}-\alpha \,|\, N_{j1} > C,\ j=1,\dots,m) = \frac{\alpha^2(v_1-\mu_1-\mu_1^2\alpha)}{(1+\mu_1\alpha)^2} \quad (2.39)$$

2.2.4 Discussion

There are some interesting findings that can be discussed based on the work of Cook and Wei (2002). First, although selection effects may be reduced for some (not all) safety-effectiveness indexes by using a control-group method, finding enough appropriate sites for the control group is much harder in traffic-safety analyses than in clinical trials. In other words, medical researchers have the luxury of setting few conditions on a patient's age, gender, or previous-response history (such as blood pressure or heart rate) to build their control group. However, locating sites with similar traffic characteristics, crash frequencies, and geographic layouts is much more difficult – especially since collecting crash and other related data is expensive and time-consuming. For instance, the study period for analyzing traffic countermeasures is usually much longer than that used for clinical trials, in part because the countermeasures typically need to have been built prior to the analysis. Additionally a significant amount of time is needed to collect a pool of data that is large enough - especially in the *after* period - to make proper inferences as to the countermeasures' effectiveness. Furthermore, there are many related factors that may change during the experiment on traffic safety. Transportation safety analysts have less control over experimental subjects (e.g., sites, road segments), as they cannot control for the behavior of drivers who travel through the areas that are part of the study. Based on the above reasons, using a control group to reduce site-selection effects is not as practical as those used in clinical trials.

The second problem with Cook and Wei's results is the true values for the parameters, $\rho, \sigma^2, \phi, \mu,$ and α are seldom known. To apply their equation, the following equations and equation (2.37) are used as their estimators:

$$\hat{\rho} = Corr(Y_{i1}, Y_{i2} | Y_{i1} > C) \tag{2.40}$$

$$\hat{\sigma}^2 + \hat{\phi} = Var(Y_{i2} | Y_{i1} > C) \tag{2.41}$$

$$\hat{\mu}_1 = E(Y_{i1} | Y_{i1} > C) \tag{2.42}$$

The above estimators are unbiased when C is relatively low. For more precise estimators, several methods have been previously proposed (Barnett et al., 2005; Barr and Sherrill, 1999; Cohen Jr, 1950; Formann, 2008). Equations (2.43) and (2.44) show the latest estimators from (Formann, 2008):

$$\hat{\sigma}^2 + \hat{\phi} = Var(Y_{i1} | Y_{i1} > C) \Big/ \left(1 - \left(\frac{f(d)}{1-F(d)}\right)^2 + d\,\frac{f(d)}{1-F(d)}\right) \tag{2.43}$$

$$\hat{\sigma}^2 = \frac{Var(Y_{i1} - Y_{i2} | Y_{i1} > C)}{2} \tag{2.44}$$

However, equation (2.43) includes a variable, d, which is a function of $\hat{\mu}$ and $\hat{\sigma}$, so equation (2.43) has to be solved iteratively. Because of this, this study still used equations (2.37), (2.40), (2.41) and (2.42) for evaluating the simulation efficiency. For more details about how to solve equations (2.43) and (2.44), the reader is referred to Formann (2008).

Low sample size is an additional problem that may affect the accuracy of the estimating equations (2.30), (2.31) and (2.39), as the aforementioned equations are asymptotically unbiased. This study will examine the appropriateness of these equations when the sample size is small.

The final problem to be resolved is the different boundary conditions for θ and δ in count and continuous data. Previous studies have reported that there was no RTM or site-selection bias when the correlation coefficient was one, but the above statement is not true as shown in equations (2.9), (2.11), (2.34), and (2.39). However, Cook and Wei did not extend their results to figure out what occurs when site-selection bias becomes a minimum value. Hence, this study will define the boundary condition by applying the estimators of mean and variance in the truncated model from Geyer (2007). This study focuses on the index of effectiveness (θ) for count data and difference (δ) for continuous data, because most traffic studies use the above indexes to show the effects of traffic-safety countermeasures. However, site-selection bias associated with the difference and the dispersion parameter for count data will be discussed below. For further discussion of the first index, we simplified equation (2.31) by dividing its denominator and numerator by μ_1:

$$\lim_{m\to\infty} E\left\{\hat{\theta}-\theta \mid N_{i1} > c_c, i=1,\cdots,m\right\} = -\theta\left[\frac{\mu_1-\Lambda_1}{\mu_1(1+\Lambda_1\alpha)}\right] = \theta\left[\frac{\Lambda_1/\mu_1 - 1}{(1+\Lambda_1\alpha)}\right] \tag{2.45}$$

With equation (2.45), it becomes more obvious that setting higher criteria will cause a larger bias because the truncated expected value (μ_1) increases. Larger values of the index (θ) will also increase the bias. It should be noted that μ_1 is the function of entry criteria (C), a dispersion parameter (α), and a mean response rate in the *before* period (Λ_1) (Geyer, 2007). Simulated data will be used to confirm this finding in the following section (Section 4).

The site-selection bias will become zero when $\mu_1 = \Lambda_1$, as shown in equation (2.45). In other words, site-selection bias will always exist until C<0, even when $\rho = 0$ (i.e., $\alpha = 0$). Thus, when C is a nonnegative integer (0, 1, 2, …), the site-selection bias will constantly be present. This also means that for crash data, this bias will exist even when we include sites with a minimum of one single crash (as long as $\alpha \neq \infty$). It should be pointed out that for count data, the site-selection effect will cause different biases for different estimators: $\Lambda_1, \Lambda_2, \theta$, and δ. Some of these will be discussed in subsequent sections.

2.3 Truncated Model

In Section 2.2, we made the hypothesis that researchers tend to choose experimental subjects that will generate a high response because they may: be constrained by budget and time and simply want to focus on a target population; want to easily observe significant changes within the responses obtained; wish to simply comply with warrants from manuals, such as the MUTCD (Department of Transportation et al., 2003).This approach to inquiry causes the sample distribution to be truncated from the global population distribution. In clinical studies, several researchers have discussed the application of truncated models when analyzing such datasets (Cruyff and van der Heijden, 2008; Kennedy, 2005; Lee et al., 2003). Most clinical studies have usually focused on zero-truncated models. It should be pointed out that in this study, constant-truncated models that are greater than zero will be used for the reasons described above. Geyer (2007) discussed the characteristics of certain truncated Poisson and Negative Binomial distributions. The estimators for the mean and variance are shown below:

1. Poisson Distribution:

- Mean $E_{\infty,\theta}\left\{N_1 \mid Y_1 > C\right\} = \Lambda_1 + \frac{C+1}{1+\beta}, where\ \beta = \frac{p_{\alpha,\theta}(Y_1 > C+1)}{p_{\alpha,\theta}(Y_1 = C+1)}$ (2.46)

- Variance: $\frac{\partial E_{\infty,\theta}\left\{N_1, N_1 > C\right\}}{\partial\theta} = \Lambda_1\left[1 - \frac{C+1}{1+\beta}\left(1 - \frac{C+1}{\Lambda_1} \times \frac{\beta}{1+\beta}\right)\right]$ (2.47)

2. Negative Binomial distribution

- Mean: $E_{\infty,\theta}\left\{N_1, N_1 > C\right\} = \Lambda_1 + \frac{C+1}{p(1+\beta)}, where\ P = (\frac{1}{\alpha\Lambda_1 + 1})$ (2.48)

- Variance:

$$\frac{\partial E_{\infty,\theta}\left\{Y_1, Y_1 > C\right\}}{\partial\theta}$$
$$= \Lambda_1\left[1-\frac{C+1}{1+\beta}\left(1-\frac{C+1}{\Lambda_1}\times\frac{\beta}{1+\beta}\right)\right] \tag{2.49}$$
$$-\frac{C+1}{p^2(1+\beta)}\left[-(1-p)+\frac{(C+1+\alpha)(1-p)}{1+\beta}+\left(\alpha-p(C+1+\alpha)\times\frac{\beta}{1+\beta}\right)\right]$$

These equations can be used for estimating μ_1 *and* ν_1 in equations (2.30), (2.31), and (2.39). It should be noted that Gurmu and Trivedi (1992) also derived estimators of mean and variance which are very similar as equations (2.48) and (2.49) and only have slightly different. Moreover, the values estimated by using these two study methods are very close to each other based on our simulation test. Hence, we selected the equation from Geyer (2007) since it is a more recent study and its result is also supported by the simulation data.

As for the truncated continuous model, the estimators for the mean and variance are

$$E(Y_{i1}\left|C' > Y_{i1} > C\right.) = \Lambda_1 + \frac{f(\frac{C\ -\Lambda_1}{\sigma_{all}}) - f(\frac{C'-\Lambda_1}{\sigma_{all}})}{F(\frac{C'-\Lambda_1}{\sigma_{all}}) - F(\frac{C-\Lambda_1}{\sigma_{all}})}\times(\sigma_{all}) \tag{2.50}$$

$$Var(Y_{i1}\left|C' > Y_{i1} > C\right.) = (\sigma_{all}^2)\left\{1+\frac{(\frac{C\ -\Lambda_1}{\sigma_{all}})f(\frac{C\ -\Lambda_1}{\sigma_{all}}) - (\frac{C'-\Lambda_1}{\sigma_{all}})f(\frac{C'-\Lambda_1}{\sigma_{all}})}{F(\frac{C'-\Lambda_1}{\sigma_{all}}) - F(\frac{C-\Lambda_1}{\sigma_{all}})} - \left[\frac{f(\frac{C\ -\Lambda_1}{\sigma_{all}}) - f(\frac{C'-\Lambda_1}{\sigma_{all}})}{F(\frac{C'-\Lambda_1}{\sigma_{all}}) - F(\frac{C-\Lambda_1}{\sigma_{all}})}\right]^2\right\} \tag{2.51}$$

where $\sigma_{all}^2 = \phi + \sigma^2$

These equations can be used for estimating μ_1 and ν_1 in equations (2.41) and (2.42). It should be noted that this study used the negative binomial model for count data. Also, equations (2.41) and (2.42) are consistent with Hauer (1980a; 1980b) results. Hauer’s results show that higher entry criteria, and a shorter observation duration may cause higher bias. His results are consistent with our assumptions. Although the value of site-selection bias could be estimated using the above equations, the variation linked to this method produces very large values especially when the sample size is small.

2.4 Hot Spot Identification

Before evaluating different countermeasures, there is always a need to identify sites where these countermeasures will be implemented. In an ideal world, the countermeasures should be randomly applied to different sites (e.g., intersections, sites, etc.). This approach is similar to studies done in medical trials. However, in practice, and by virtue of the characteristics of the sites, the countermeasures are nearly always installed at sites that experience larger-than-expected crash counts. This means that the identification of sites can also be influenced by what entry criteria are used.

Erroneously selecting sites for applying countermeasures not only wastes limited funds that are allocated for such purposes, but it may also incur major social costs, as truly dangerous sites might go unidentified and thus crashes there will continue. Many research studies have been conducted over the last 30 years about different methods of identifying hot spots or block zones (e.g., see (Geurts and Wets, 2003; Miranda-Moreno, 2006). These selection thresholds may help traffic engineers and transportation-safety analysts focus on high-risk crash locations; however, it may also cause bias in evaluating the effectiveness of the countermeasures, because the sample distribution may not be the same as the population distribution.

According Miranda-Moreno (2006), the most common methods of identifying hot spots can be broken up into two types of strategies: threshold-based, and budget-limit. The threshold-based strategy is used to identify a list of hazardous sites whose crash counts exceed a certain value. This value may be a fixed number or a probability – for example, all sites having more than three crashes per year, or those that experience a high likelihood (i.e., > 80%) of having five or more crashes per year. This strategy ensures that the selected locations on the list are considered dangerous at some critical level, with the number of locations to be selected unspecified. This is the most appropriate strategy when applying local safety policies that identify tolerance levels for accident risks. The shortcoming with this strategy is related to how one formally defines the thresholds or decision rules for generating the optimal selection of a hotspot list (Higle et al., 1988; Schlüter et al., 1997). Different hotspot priorities have been proposed by several researchers ((Hauer and Persaud, 1987; Heydecker and Wu, 2001; Miaou and Song, 2005; Persaud et al., 1999; Schlüter et al., 1997). They include the expectation of accident frequency, probability of excess, potential of accident reduction, and expectation of ranks.

The other aforementioned strategy for identifying hotspots, budget-limit, consists of identifying a list of hazardous sites using available budget and crash costs. This study will not evaluate this strategy.

Practically speaking, this study will compare the possible bias in estimating the effectiveness when using the fixed-number approach for setting entry criteria in a trial. This study will also focus on the change of bias in different scenarios, such as high/low crash rates, high/low heterogeneity, and high/low entry criteria levels.

2.5 Methods for Estimating the Effectiveness of Countermeasures

In the literature, several methods have been proposed for estimating the effectiveness of countermeasures. For count data, they include the naïve before-after study, the C-G method, and the EB method. For continuous data, they include the naïve before-after study, the C-G method, and the ANCOVA method. Additional information about these three methods can be found in (Barnett et al., 2005; Hauer, 1997). Recently, a more advanced method, known as the Full Bayes (or FB) method, has also been applied to before-after studies (Park et al., 2010). So far, these methods, with the exception of the naïve before-after study, have been used for minimizing the effects attributed to RTM. Very few studies have focused on using site-selection effects with these different methods in order to estimate the effectiveness of particular countermeasures. As such, this proposed study will evaluate how site-selection effects can be incorporated into the above methods of estimating the effectiveness of countermeasures. Moreover, this study has also proposed a new adjusted method of removing site-selection effects by extending Cook and Wei's results. Then, how to remove site-selection bias by using FB method will be discussed in the third subsection in the methodology section.

2.6 Other Methods of Minimizing Site-Selection Bias

Unlike our adjusted method of estimating site-election bias and removing it from the naïve effectiveness estimator, another type of method is to reduce the site-selection bias itself. The following paragraphs will introduce these methods individually.

Aside from the control group method that was mentioned earlier, other common methods for reducing site-selection bias include: increasing the baseline measurement, and the repeat screen test (Barnett et al., 2005; Davis, 1976; Yudkin and Stratton, 1996). Detailed information of the above-mentioned methods is introduced below.

1. Increasing the baseline measurement: Taking the average of a number of measurements as the baseline measurement can reduce the RTM by decreasing the variability of the criterion measurement. Equation (2.52) shows that the site-selection bias decreases when the baseline measurement increases to k. Please see (Davis, 1976) for details.

$$\delta-\hat{\delta}=\frac{\sigma^2/k}{\sqrt{\sigma^2/k+\phi}}\frac{f(d)}{(1-F(d))} \tag{2.52}$$

2. Repeat screen test: Use the first observation to catalog subjects and the second and third observations to calculate treatment effectiveness; the RTM will be removed while the correlation between the first and second observations equals to the correlation between the first and third observations. Please refer to the Davis, (1976) and Ederer (1972) studies for more-detailed information.

3. Analysis of Covariance (ANCOVA): The ANCOVA has been utilized extensively in epidemiology research (Barnett et al., 2005). Basically, the ANCOVA provides a more precise estimate of treatment effect by adjusting each subject's follow-up measurement according to its baseline measurement. The characteristics of this method are described in the methodology section further below. Although the ANCOVA has successfully been used in the past in the context of a before-after study with entry criteria (i.e., its estimator has a narrower confidence interval than the traditional paired t-test), the method has the same disadvantage as the one used for the CG method: both methods require collecting additional data that will be part of the control group, which may be prohibitive depending on the study characteristics. Additional information can be found in Barnett et al.(2005), who applied the ANCOVA for analyzing the effects of skin cancer prevention treatments in Nambour, Australia.

2.7 Summary

Below are the three primary findings based on the previous literature review:

1. Site-selection effects exist when entry criteria are used for selecting observations, and this selection process is very common in traffic safety.
2. Few studies have examined how the magnitude of a site-selection effect is influenced by setting different entry criteria, with and without the use of a control group (the use of a control group is not as practical in highway safety).
3. Many methods have been proposed for estimating the effectiveness of countermeasures, but few have considered the relationship between entry criteria and site-selection bias.

In sum, the purpose of this study is to develop a new method that can easily adjust or reduce naïve estimators in before-after studies that employ continuous data and that are characterized by an entry criterion, without relying on data collected for the control group. This will be performed by comparing the new method (subsequently called the

Adjusted Method in the text) with the other common before-after methods: the Naïve, CG, and EB (for count data)/ANCOVA (for continuous data). For practical purposes, this study will compare possible bias when estimating effectiveness based upon the methodologies used for identifying hot spots by fixed-number. This study will also focus on the change of bias in different scenarios, such as high/low crash rates, high/low heterogeneity, and high/low entry criteria. The next section describes the methodologies regarding how to examine the site selection bias for continuous data and count data.

3. METHODOLOGY

This section describes the methodologies regarding how to examine the site selection bias in two areas: 1) Estimating countermeasure effectiveness for continuous data; 2) Estimating countermeasure effectiveness for count data. Each section focuses on the bias caused by using common methods in before-after studies for different types of data and their safety index respectively.

3.1 Methods to Examine Countermeasure Effectiveness for Continuous Data

Section 3.1 includes four before-after methods for normal-distribution data: Naïve before-after studies, Before-after studies with a control group (CG), ANCOVA, and our adjusted method. The first three methods are the most common types of before-after studies, and the last one is a new method that we have proposed in this study. The idea of this new method is to remove/reduce site-selection bias from the naïve estimator. The bias estimating equation has also been updated from (Cook and Wei, 2002). The difference, δ, was chosen as the index of countermeasure effectiveness for continuous data, because it is commonly used. This study only assumed one year for the *before* and *after* periods, respectively, in order to simplify the calculations.

NOTATIONS

Before describing the methodology, it is important to define the notations used in the study:

C, C': The entry criterion (minimum and maximum);

m: The sample size;

$\Lambda_{i1}^{T}, \Lambda_{i1}^{CG}$: The mean response rate for site i (T: treatment group, CG: control group) in the before period (k=1);

$\Lambda_{i2}^{T}, \Lambda_{i2}^{CG}$: The mean response rate for site i (T: treatment group, CG: control group) in the after period (k=2);

Y_{i1}^{T}, Y_{i1}^{CG}: The observed response for site i (T: treatment group, CG: control group) in the before period for continuous data, $Y_{i1}^{T}, Y_{i1}^{CG} >$ C;

Y_{i2}^{T}, Y_{i2}^{CG}: The observed response for site i (T: treatment group, CG: control group) in the after period for continuous data;

$\hat{\delta}_{ANCOVA}$: The estimator of difference by using the ANCOVA method; this is the coefficient (b) of the "group" variable in the regression model; and,

$$Y_{i2} = \text{constant} + a \times (Y_{i1} - \overline{Y}_{i1}) + b \times group + error\,,$$

Where group=1, treatment group; group=0, control group.

Given the notation above, it is now possible to define the equations of difference, δ, which are estimated by four before-after methods for one-sided truncated normal distribution as below:

(1). Naïve method: $\hat{\delta}_{naive} = (\hat{\Lambda}_2^T - \hat{\Lambda}_1^T) = \dfrac{\sum_{i=1}^{m} Y_{i2}^T - \sum_{i=1}^{m} Y_{i1}^T}{m}$ (3.1)

(2). CG method:

$$\begin{aligned}\hat{\delta}_{CG} &= (\hat{\Lambda}_2^T - \hat{\Lambda}_1^T) - (\hat{\Lambda}_2^{CG} - \hat{\Lambda}_1^{CG}) \\ &= \frac{1}{m}\left[(\sum_{i=1}^{m} Y_{i2}^T - \sum_{i=1}^{m} Y_{i1}^T) - (\sum_{i=1}^{m} Y_{i2}^{CG} - \sum_{i=1}^{m} Y_{i1}^{CG}) \right]\end{aligned} \tag{3.2}$$

(3). ANCOVA method: $\hat{\delta}_{ANCOVA} = \hat{b}$ (3.3)

(note: b was described above equation (2.47))

(4). Adjusted method:

$$\hat{\delta}_{adjusted} = \hat{\delta}_{naive} - (\hat{\rho} - 1)\sqrt{(\hat{\phi} + \hat{\sigma}^2)}\, f(\hat{d}) \Big/ (1 - F(\hat{d}))$$

$$= \hat{\delta}_{naive} - (Corr(Y_{i1}^T, Y_{i2}^T) - 1)\sqrt{Var(Y_{i2}^T)} \left[f\left(\frac{C - \bar{Y}_1^T}{\sqrt{Var(Y_{i2}^T)}}\right) \Big/ \left(1 - F\left(\frac{C - \bar{Y}_1^T}{\sqrt{Var(Y_{i2}^T)}}\right)\right) \right] \tag{3.4}$$

3.2 Methods to Examine Countermeasure Effectiveness for Count Data

In the same way as above, Section 3.2 also includes four methods for count data: Naïve before-after studies, Before-after studies with a control group, Empirical Bayes, and our adjusted method. For the sake of convenience, the common notations are listed below:

θ: The safety effectiveness, $0 \leq \theta \leq 1$ (could be higher, but not in this study)

C: Entry criteria

n: Sample size

α: Dispersion parameter

$\Lambda_{i1}^T, \Lambda_{i1}^C$: Mean response rate for site i (T: treatment group, C: control group) in the before period, i=1, …, n

$\Lambda_{i2}^T, \Lambda_{i2}^C$: Mean response rate for site i (T: treatment group, C: control group) in the after period, i=1, …, n

N_{ij1}^T, N_{ij1}^C: The observed response for site i (T: treatment group, C: control group) in j year (in the before period) for count data, $N_{ij1}^T > C$

N_{ij2}^{T}, N_{ij2}^{C}: The observed response for site i (T: treatment group, C: control group) in j year (in the after period) for count data

W_i: Weight for site i in EB method, $W_i = \dfrac{1}{1+\hat{\Lambda}_{i1}^{T} \times \sum_{j=1}^{t} N_{ij1} \times \alpha}$

Λ_1: the estimator for the average crash rate of all sites in the before period,

$\Lambda_1 = E(N_{i1} | N_{i1} > C) = \mu_1$ for the EB$_{MM}$ method, and $\Lambda_1 = \Lambda_1$ for the EB$_{CG}$ method.

M_{i1} : Expected responses for site i in EB method,

$$M_{i1} = W_i \times (\Lambda_{i1}) + (1 - W_i) \times (\sum_{j=1}^{t} N_{ij1})$$

Given the notation above, it is now possible to define the equations of safety index, θ, which is estimated by four before-after methods for a one-sided truncated count distribution as below:

$$\text{(1). Naïve method: } \hat{\theta}_{naive} = \frac{\sum_{i=1}^{n} \hat{\lambda}_{i2}}{\sum_{i=1}^{n} \hat{\lambda}_{i1}} = \frac{\frac{1}{n}\frac{1}{t}\sum_{i=1}^{n}\sum_{j=1}^{t} N_{ij2}^{T}}{\frac{1}{n}\frac{1}{t}\sum_{i=1}^{n}\sum_{j=1}^{t} N_{ij1}^{T}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{t} N_{ij2}^{T}}{\sum_{i=1}^{n}\sum_{j=1}^{t} N_{ij1}^{T}} \tag{3.5}$$

$$\text{(2). CG method: } \hat{\theta}_{CG} = \frac{\hat{\lambda}_{2}^{T}}{\hat{\lambda}_{2}^{T} \times \frac{\hat{\lambda}_{2}^{C}}{\hat{\lambda}_{1}^{C}}} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{t} N_{ij2}^{T}}{\sum_{i=1}^{n}\sum_{j=1}^{t} N_{ij1}^{T} \times \sum_{i=1}^{n}\sum_{j=1}^{t} \frac{N_{ij2}^{C}}{N_{ij1}^{C}}} \tag{3.6}$$

(3). Empirical Bayes: $$\hat{\theta} = \frac{\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{t} N_{ij2}^{T}}{\sum_{i=1}^{n}\sum_{j=1}^{t} M_{ij1}^{T}} \right)}{\left\{ 1 + \left[\frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{t} \mathrm{Var}\left(M_{ij1}^{T}\right) \right)}{\left(\sum_{i=1}^{n}\sum_{j=1}^{t} M_{ij1}^{T} \right)^{2}} \right] \right\}} \quad (3.7)$$

(Note: the denominator is used to adjust for the small sample size)

(4). Adjusted method:

$$\hat{\theta}_{adjusted} = \hat{\theta}_{naive} + \theta \left[\frac{\mu_1 - \Lambda_1}{\mu_1(\Lambda_1 \alpha + 1)} \right]$$

$$= \hat{\theta}_{naive} \left[1 + \frac{C+1 \Big/ (1 + \frac{P(N > C+1)}{P(N = C+1)})}{\hat{\Lambda}_1 + \left[\frac{C+1}{(\hat{\Lambda}_1 \hat{\alpha} + 1)^{-1} (1 + \frac{P(N > C+1)}{P(N = C+1)})} \right]} \right] \quad (3.8)$$

Where $$\hat{\Lambda}_1 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{t} N_{ij1}^{T} \quad (3.9)$$

$$\alpha = \frac{\left[\frac{E((N_{i2} - \Lambda_1)^2 \,|\, N_{i1} > C)}{E(N_{i2} \,|\, N_{i1} > C)} - 1 \right]}{E(N_{i2} \,|\, N_{i1} > C)} \quad (3.10)$$

Equation (3.8) is based on two assumptions:

$\mu_1 = \Lambda_1 + \left[\frac{C+1}{(\hat{\Lambda}_1 \hat{\alpha} + 1)^{-1} (1 + \frac{P(N > C+1)}{P(N = C+1)})} \right]$ and $\hat{\theta} = \hat{\theta}_{naive}$.

Also, according to equation (2.39) the adjusted estimator of dispersion parameter, α , is defined as:

$$\hat{\alpha}_{adjusted} = \hat{\alpha}_{naive} + \hat{\alpha}_{naive}^{2}\left[\frac{v_1 - \mu_1 - \mu_1^{\,2} v_1}{(\mu_1 \hat{\alpha}_{naive} + 1)^2}\right] \tag{3.11}$$

It should be noted that equation (3.10) was calculated by using crash data in the *after* period instead of using crash data in the *before* period which is commonly used in current before-after studies.

$$\alpha = \frac{\left[\dfrac{E((N_{i2} - \Lambda_2)^2 \,|\, N_{i1} > C)}{E(N_{i2} \,|\, N_{i1} > C)} - 1\right]}{E(N_{i2} \,|\, N_{i1} > C)}$$

$$\text{instead of } \alpha = \frac{\left[\dfrac{E((N_{i1} - \Lambda_1)^2 \,|\, N_{i1} > C)}{E(N_{i1} \,|\, N_{i1} > C)} - 1\right]}{E(N_{i1} \,|\, N_{i1} > C)} \tag{3.12}$$

Same as above, the equations of difference, δ , which is estimated by four before-after methods for a one-sided truncated count distribution as below:

(1). Naïve method: $\hat{\delta}_{naive} = \sum_{i=1}^{n} \hat{\lambda}_{i2} - \sum_{i=1}^{n} \hat{\lambda}_{i1} = \frac{1}{n}\frac{1}{t}(\sum_{i=1}^{n} \sum_{j=1}^{t} N_{ij2}^{T} - \sum_{i=1}^{n} \sum_{j=1}^{t} N_{ij1}^{T})$ (3.13)

(2). CG method: $\hat{\delta}_{CG} = \hat{\lambda}_2^T - \hat{\lambda}_1^T \times \frac{\hat{\lambda}_2^C}{\hat{\lambda}_1^C} = \frac{1}{nt}(\sum_{i=1}^{n} \sum_{j=1}^{t} N_{ij2}^{T} - \sum_{i=1}^{n} \sum_{j=1}^{t} N_{ij1}^{T} \times \sum_{i=1}^{n} \sum_{j=1}^{t} \frac{N_{ij2}^{C}}{N_{ij1}^{C}})$ (3.14)

(3). Empirical Bayes: $\hat{\theta} = \frac{1}{nt}\left(\sum_{i=1}^{n} \sum_{j=1}^{t} N_{ij2}^{T} - \sum_{i=1}^{n} \sum_{j=1}^{t} M_{ij1}^{T}\right)$ (3.15)

(4). Adjusted method:

$$\hat{\delta}_{adjusted} = \hat{\delta}_{naive} - (\mu_1 - \Lambda_1)\frac{(\delta\alpha - 1)}{(\Lambda_1\alpha + 1)}$$
$$= \hat{\delta}_{naive} - \left[\frac{(C+1)\times(\hat{\delta}_{naive}\hat{\alpha} - 1)}{1 + \frac{P(N > C+1)}{P(N = C+1)}}\right] \tag{3.16}$$

3.3 Summary

In this section, the researcher provided the methodologies regarding how to examine the site selection bias in two areas: 1) Estimating countermeasure effectiveness for continuous data; 2) Estimating countermeasure effectiveness for count data. There are different distribution model to fit traffic safety data, but this section we have focused on the normal distribution for continuous data and Negative Binomial distribution for count data.

Each section focused on the bias caused by using common methods in before-after studies for different types of data and their safety index respectively. This section also explains how the Adjusted method by combing Cook and Wei's (2002) and Geyer's (2007) results were derived. Also, prior to describing these before-after methods, detail notations have been provided. Each section focuses on the bias caused by using common methods in before-after studies for different types of data and their safety index respectively. Based on these equations, the next three sections apply the methodology to simulated dataset (section 4) and observed datasets (section 5 and 6).

4. QUALIFICATION OF SITE-SELECTION BIAS

In this section, simulated data were generated and used to support the above theorem and to examine the accuracy of the previous estimates of site-selection bias, because these equations are modified from Cook and Wei (2002)[1]. Sections 4.1 and 4.2 provided possible factors, simulation protocol, and simulation results of the site-selection bias for continuous and count data respectively. In Section 4.3, count data simulated from vary mean (instead of a fixed mean in Section 4.2) were used in order to make sure the above finding also adapted in practical. Section 4.4 summarized the main findings in this section.

4.1 Continuous Data

This section consists of two parts. The first part describes the methodology and simulation protocol used for estimating the bias for five scenarios. The second part provides simulation results by scenarios. This study only assumed one year for the *before* and *after* periods, respectively, in order to simplify the calculations.

4.1.1 Scenarios for Possible Factors

In the following subsections four possible factors of influence will be discussed by sensitive analysis: different before-after methods, different between-subject variances, different within-subject variances, and different sample sizes. Note that the estimators were estimated the same way for the subsequent scenarios. For subsequent scenarios, the following variables were assumed to be fixed: sample size equal to 100 (except scenario

---

[1] Although theoretical equations to estimate the site-selection bias are available, simulation is needed for three reasons: (1) the theoretical derivations proposed by Cook and Wei (2002) had to be modified because it included several typos, and the simulation results are used to verify the accuracy of these modified equations; (2) the theoretical derivations assumed that the sample is infinite, which does not reflect how these equations would be applied using observed data; (3) and, simulation was also used to examine dissimilar control groups and the accuracy of the bias-estimation equations when the parameters are unknown.

4) and the difference equal to 10. Note that scenarios 1 and 2 were analyzed simultaneously.

Scenario 1: Direct Comparison of the Methods

This scenario assumed that the entry criteria varied from C=58 (i.e., $Y_{i1} > 58$) to C=73, and that the mean value was 70. We can assume, for instance, that the average speed within various segments is 70 mph and the minimum driving speed limit is 58, 61, …, or 73. The data-generating procedure followed a normal distribution to generate the between-subject variance, within-subject variance, and observed speed. m (i=1 to m) observations were selected randomly for the treatment group only when the response was larger than the entry criteria ($Y_{i1} > C$). We labeled responses of these sites as Y_{i1}^T and Y_{i2}^T. The estimators for the measure of difference are as above in equations (3.1), (3.2), (3.3), and (3.4).

Recall that the estimators of $\rho, \sigma^2, \phi,$ and μ in equation (3.4) are based on equations (2.40) to (2.42) when their true values are unknown. It should be noted that the estimators were calculated in the same manner for the subsequent scenarios. For this scenario, the following variables were assumed to be fixed: sample size equal to 100, between-subject variance equal to 25 (=5^2), and a safety difference equal to 10. Note that Scenarios 1 and 2 were analyzed simultaneously.

Scenario 2: Between-subject variance (ϕ)

As discussed in previous studies (Barnett et al., 2005; Cook and Wei, 2002), a higher correlation coefficient reduces the site-selection bias for estimating the difference, but it also creates a higher bias for estimating the mean rate for the *after* period. Also, the correlation coefficient is a function of the between-subject variance and the within-subject variance. Speed measurements can be used as an example. High between-subject variance means that speed varies significantly among individual drivers. High within-subject variance means that the speeds for the same driver in the before and after periods are very different. To better understand the extent of this bias, Scenarios 2 and 3 examined the impacts on the selection bias for these two variances separately. Scenario 2 assumed that the between-subject variance varied from 9 (=3^2, small heterogeneity) to 225 (=15^2, very large heterogeneity). There are seven levels of variance: 3^2, 4^2, 5^2, 7^2, 9^2, 11^2, and 15^2. It should be pointed out that the between-subject variances that have been observed with real speed data are from 4^2 to 7^2 (Muchuruza and Mussa, 2004).

Scenario 3: Within-Subject Variance (σ^2)

This scenario assumed that the within-subject variance varied from 9 (small heterogeneity), 16 (medium heterogeneity), and to 25 (large heterogeneity). For this scenario, the following variables were assumed to be fixed: a sample size equal to 100 and the difference equal to 10.

Scenario4: Sample Size

Equation (2.8) was derived with the assumption that sample size tends towards infinity. There is a need to examine whether the estimator changes or not when sample size is reduced. The scenario assumed that the sample size varied as follows: 10 (small), 30 (medium), and 100 (large). For this scenario, the difference was equal to 10.

4.1.2 Simulation Protocol

The simulated data were generated using the software R (R Development Core Team, 2006)). The general steps were as follows:

(1). For each between-subject variance (ϕ), generate the subject-specific random effect (u_i), which follows a normal distribution $N(0,\phi)$. The sample mean in the *before* period is 70, and error term of each site also follows a normal distribution $N(0,\sigma^2)$. σ is equal to 5 for all scenarios except for Scenario 3, which examines the effects of different σ values. Then, observed data (Y_i) in the before period were generated by combining the above terms: mean, random effect, and error for each site i ($Y_{ik} = \mu_k + u_i + e_{ik}$).

(2). Observed data in the after period were generated using a similar procedure, and the only difference was that the mean in the *after* period was equal to 80.

(3). Generate the data for 5,000 sites, but randomly select 100, 30 and 10 sites depending on the scenario.

(4). Then, m sites are selected as the sample, whose observed (speed) values are larger than the entry criteria (58, 61,…, 73). The effectiveness can be estimated using equations (3.1), (3.2), (3.3), and (3.4).

(5). When the control group is used, $\delta_{control}$ is equal to 0. In other words, there is no difference in the mean rate between the *before* and *after* periods for the control group.

(6). Repeat Steps 2 to 5 for a total of 1,000 times, and estimate the biases of various estimators $\left(\hat{\delta}_{1000} - \delta\right)$.

4.1.3 Simulation Results

This section describes the results based on the above simulation protocol. The results are presented for each scenario. As mention earlier, four possible factors include different before-after methods, different between-subject variances, different within-subject variances, and different sample sizes.

Scenario 1 - RESULTS

Figure 4-1 shows the site-selection bias for the Naïve, CG, ANCOVA, and the Adjusted methods. Overall, this figure shows that the bias is reduced as the between-subject variance increases, except when C is particularly small (e.g. C=58). This was expected given the characteristics of equations (2.8) and (3.4). The greater the entry criteria, the more biased the estimate will be. Among the four methods, the Naïve method (Figure 4-1(a)) has highest site selection bias; δ can be over-estimated by as much as 49%. As discussed by Cook and Wei (2002), unless the correlation coefficient is close to one (which means $\phi >> \sigma^2$), $\hat{\delta}$ will be biased if an entry criterion is used (e.g. the bias never equals zero even when $\phi = 225$).

When the CG or ANCOVA method is used, the bias can be theoretically eliminated. For the CG method (Figure 4-2(b)), the control group needs to have the same characteristics (i.e., the same sample mean, variance, and entry criteria) as those of the treatment group used for the Naïve method. As explained above, it may be difficult in practice to find datasets with the exact same characteristics as those of the treatment group.

The application of the control group is further explored in Figure 4-2. Figure 4-2(a) is the same as Figure 4-1(b), and is used to compare the results with the other figure below. Figure 4-2(b) shows that when the control group has a slightly lower mean (0.95 × mean), the site-selection bias is far from zero but smaller than the bias generated from the Naïve method. Furthermore, the bias is in the opposite direction (CG with a lower mean will over-estimate δ). Conversely, using a control group that has a higher mean value (1.2 × mean) causes a negative site-selection bias, which is even higher than with the Naïve method (Figure 4-2(c)). In other words, the site-selection bias caused by using a dissimilar control group might be even higher than with using just the Naïve method. Because of space limitations, other figures showing site-selection bias for the ANCOVA that have a different mean are not shown here. However, the results are similar to those documented above for the CG method. Finally, using the Adjusted Method reduced site-selection bias by about 50% - even when the biased estimators of $\rho, \sigma^2, \phi,$ and μ are used Figure 4-1(d), or when they are not known with certainty. Compared to each other method, the Adjusted Method provides a more precise estimate than the Naïve method and performs better than the CG and ANCOVA methods when similar control group data are not available. It should be pointed out that the Adjusted Method will not completely eliminate the site-selection effects unless $\rho, \sigma^2,$ and ϕ are fully known. Obviously, these values are rarely known in practice.

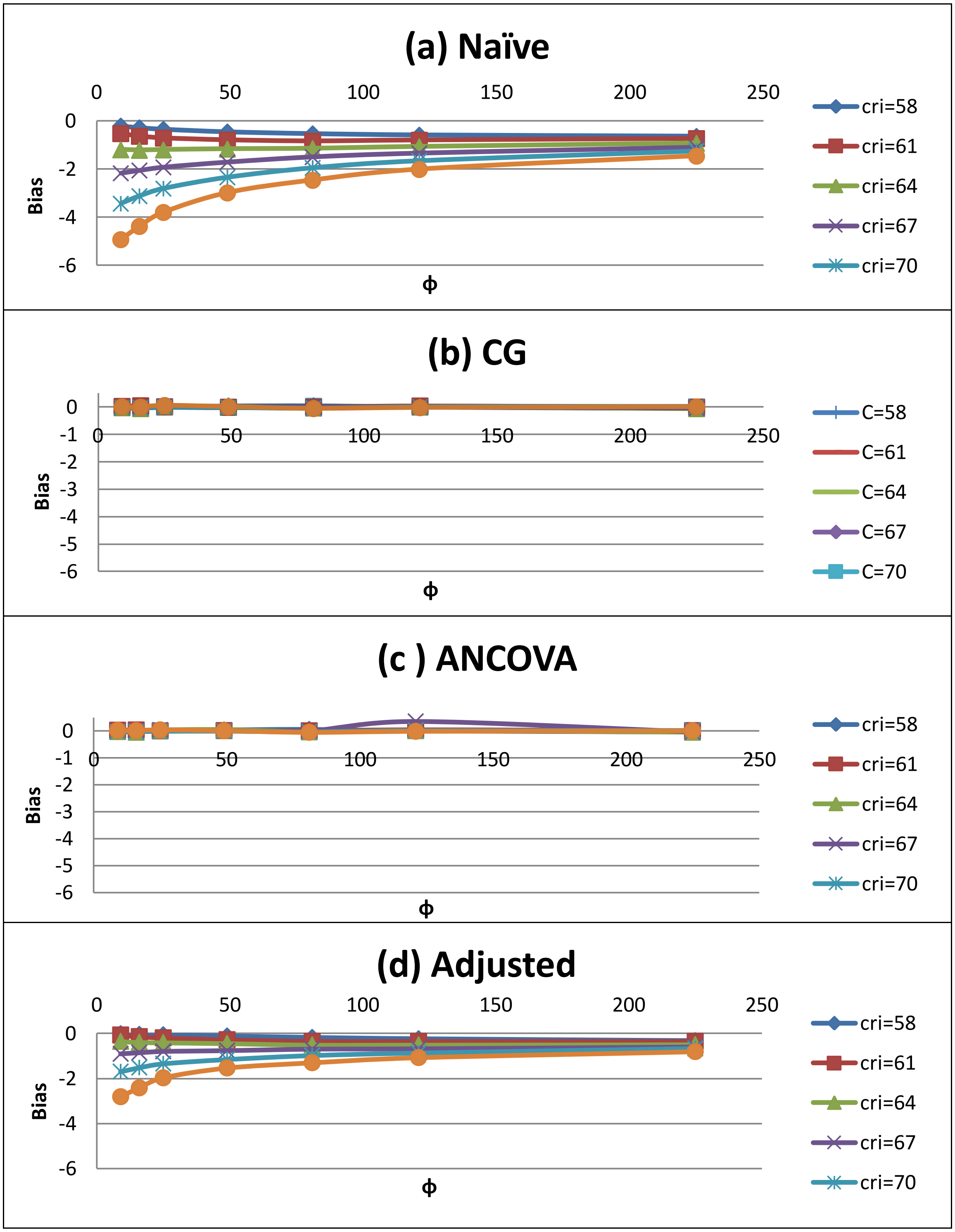


Figure 4-1 Site-selection bias for the Naïve, CG, ANCOVA, and our Adjusted methods

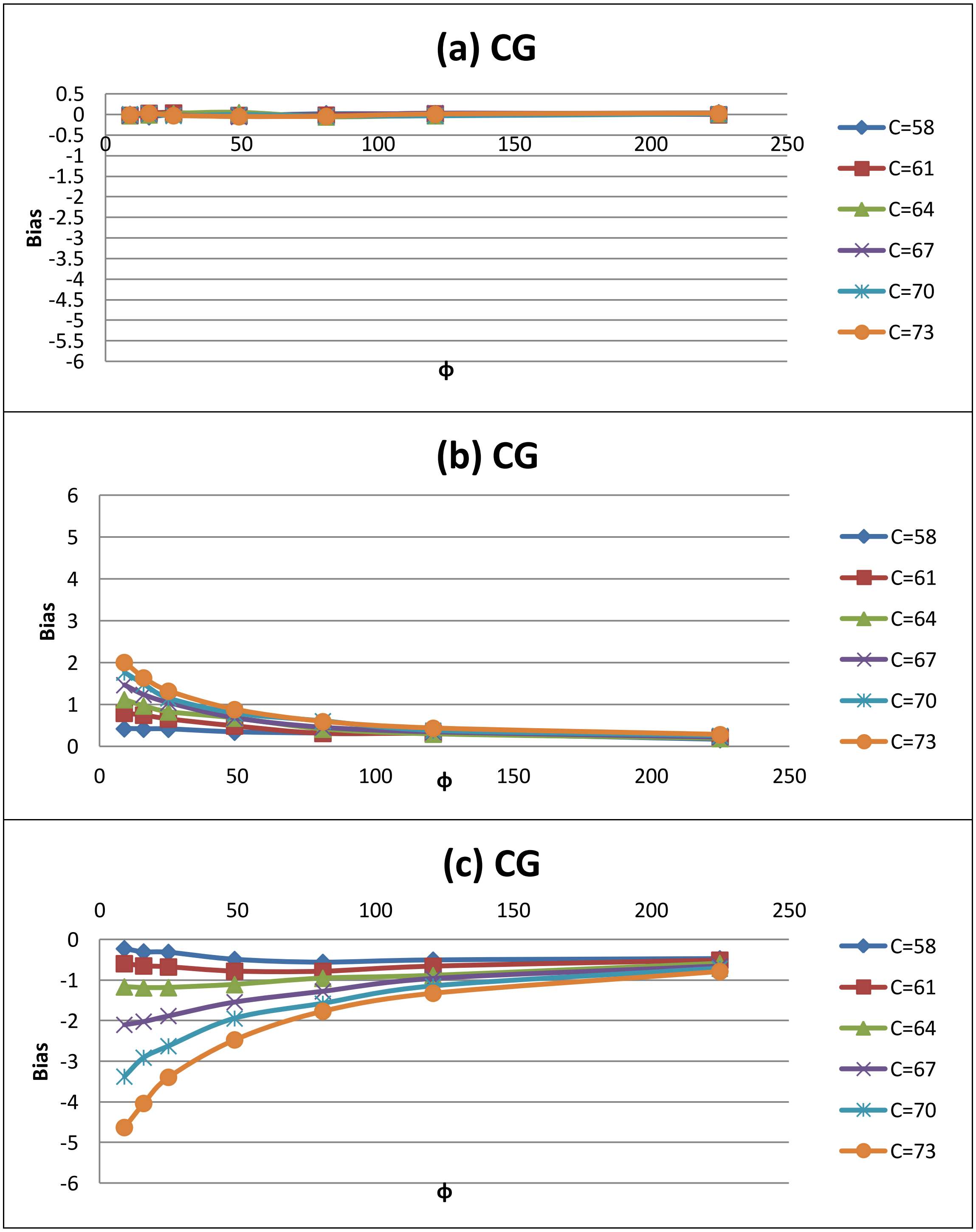


Figure 4-2 Site selection bias for the CG method for the following characteristics: (a) same mean; (b) 0.95 × mean; (c) 1.2 × mean.

Scenario 2 − RESULTS

As discussed in the first scenario, the site-selection bias decreases when the between-subject variance increases, but the decreasing rate becomes almost flat for values above 11^2 (Figure 4-1). With the Naïve method, the bias is never eliminated, as compared to the CG and ANCOVA methods. It should also be noted that when the between-subject variance tends towards zero (subjects have a low heterogeneity), the site-selection bias is the largest. This finding is consistent with Cook and Wei (2002), because the correlation coefficient is close to zero when between-subject variance is zero.

Scenario 3 – RESULTS

Figure 4-3 shows that the value of within-subject variance changes the value of the bias. This figure illustrates that the bias increases when the error term becomes larger, which is consistent with the characteristics of equation (2.8). Moreover, the Adjusted method reduces the bias by 50% for all within-subject variance. The results from Scenarios 2 and 3 indicate that lower between-subject variance and higher within-subject variance cause more selection bias.

Scenario 4 – RESULTS

Figure 4-4 shows that the sample size of the treatment group does not affect the bias considerably. For all sample sizes, the biases estimated from the Adjusted method still reduce the bias by 50%, although there is a slightly difference when sample size is small and the entry criteria is low (please see the dash line in Figure 4-4(c)). Hence, equation (2.8) and (3.4) which was derived by assuming that the sample size is close to infinity (∞), may be used for estimating site-selection biases when the sample size is small.

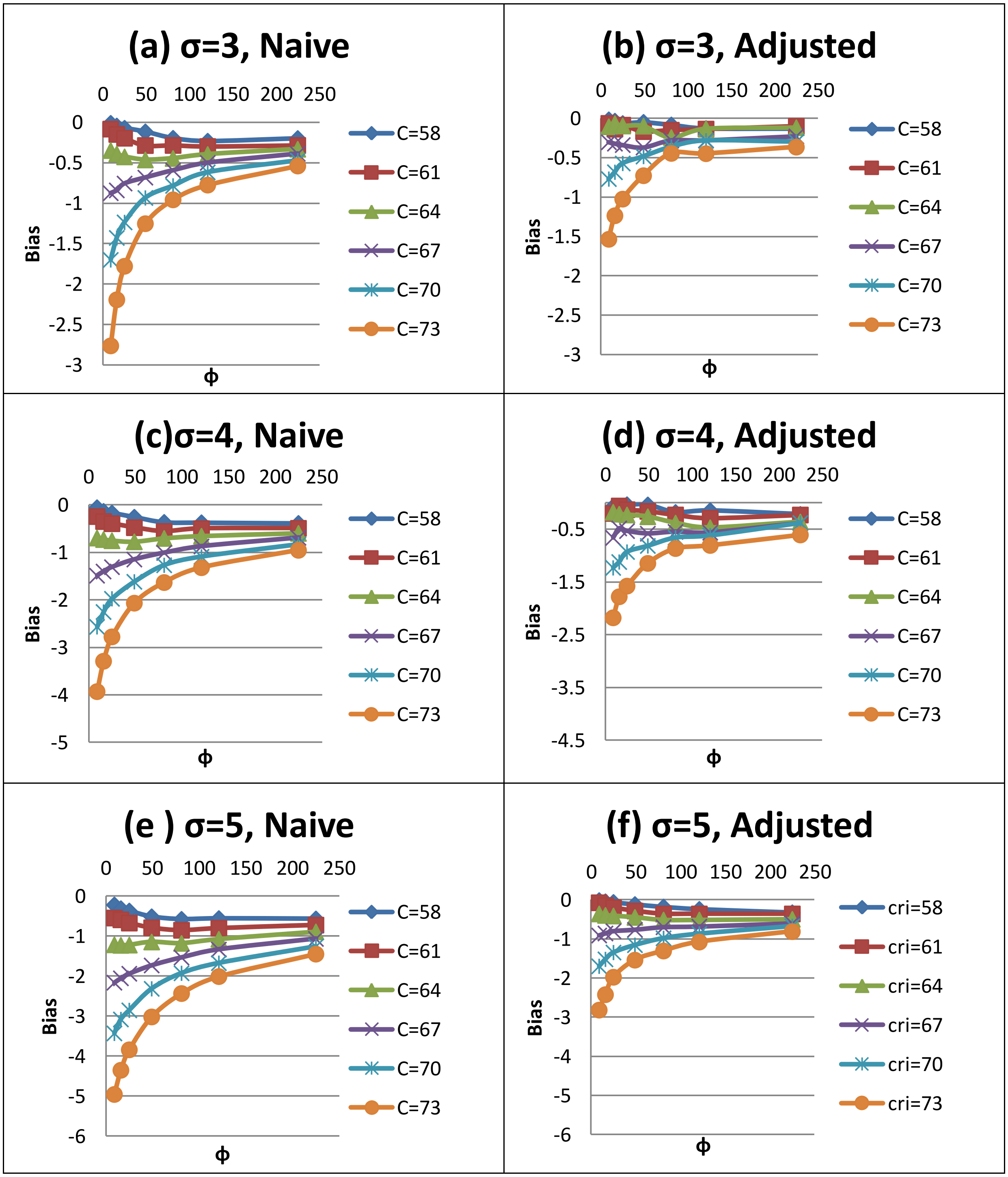


Figure 4-3 Site selection bias for the Naïve and Adjusted method when the within-subject variance is equal to 3^2, 4^2, and 5^2.

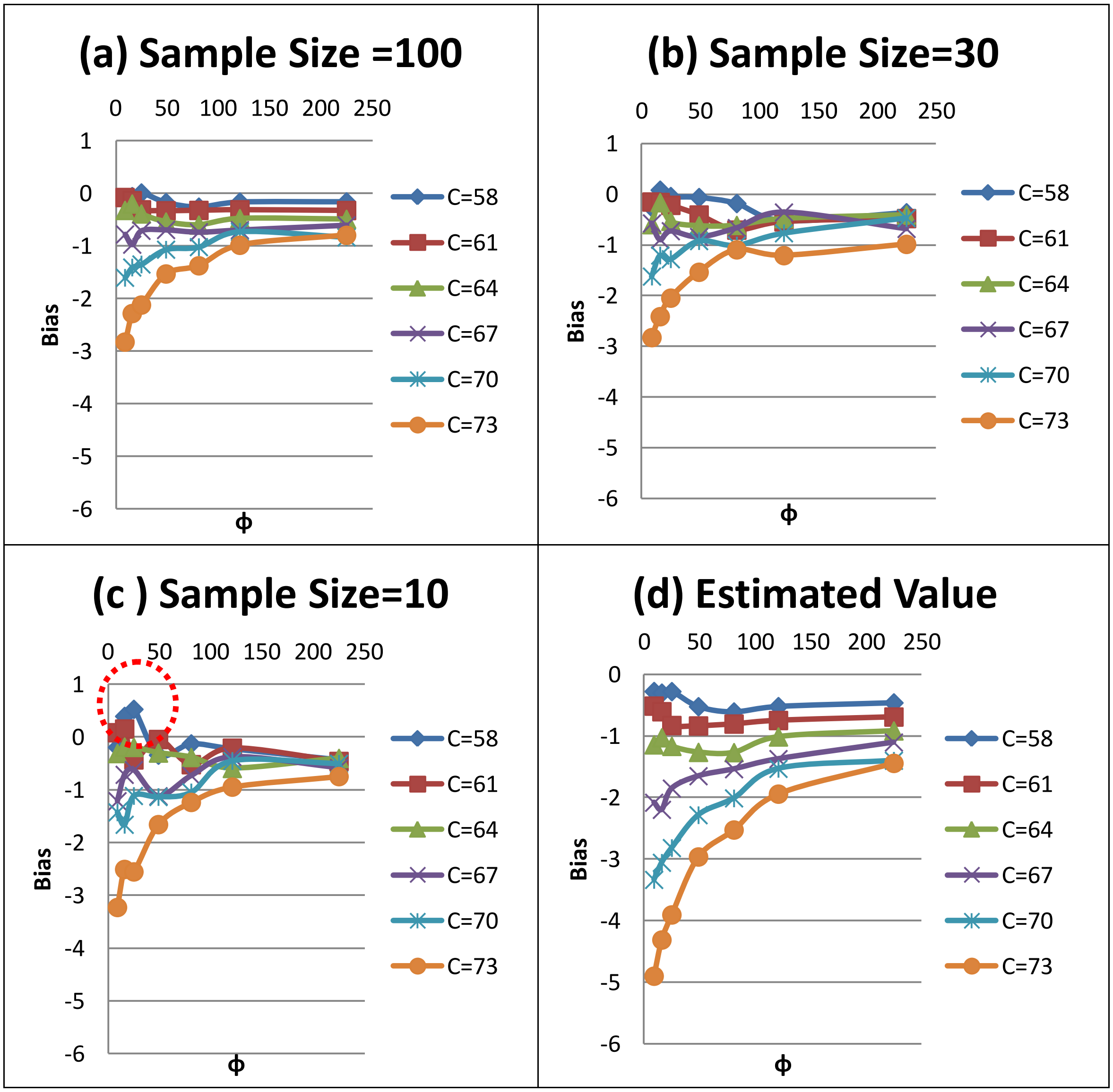


Figure 4-4 Site selection bias for the Adjusted method when the sample size is equal to (a) 100, (b) 30, (c) 10, and (d) Estimated.

4.2 Count Data: Simulation from Fixed Crash Mean

Unlike previous section only focused on one safety index (difference) for continuous data, this section examined site-selection bias on three parameters, safety effectiveness, difference, and dispersion parameter in three subsections. Each subsection consists of two parts. The first part describes the methodology and simulation protocol used for estimating the bias for five scenarios. The second part provides simulation results by scenarios.

4.2.1. Safety Effectiveness

Safety effectiveness is the most common index to measure the countermeasure effectiveness in count data, while difference is the main index in continuous data.

4.2.1.1 Scenarios for Possible Factors

In the following subsections five possible factors of influence will be discussed by sensitive analysis: entry criteria, sample size, inverse dispersion parameters, safety-effectiveness values, and the standard deviation of safety effectiveness. As mentioned above, note that the estimators were estimated in the same way as for the subsequent scenarios. For this scenario, the following variables were assumed to be fixed: sample size equal to 100 (except Scenario 2), and safety effectiveness equal to 0.50 (except Scenario 4). Note that Scenarios 1 and 2 were analyzed simultaneously.

Scenario 1: Direct Comparison of the Methods

This scenario assumed that the entry criteria varied from C=0 (i.e., $N_{i1} \geq 1$) to C=5. The data generation procedure followed a Poisson-gamma distribution to generate the mean response rate and observed number of counts. m (i=1 to m) observations were selected randomly for the treatment group only when the response was larger than the entry criteria ($\sum_{j=1}^{t} N_{ij1} > C$). We label these observes of sites as N_{ij1}^{T} and N_{ij2}^{T}. The estimators for the measure of effectiveness are the same as in equations (3.5), (3.6), (3.7), and (3.8). It should be noted that the estimators were estimated in the same way for the subsequent scenarios. For this scenario, the following variables were assumed to be fixed: sample size equal to 100, and safety effectiveness equal to 0.50. Note that Scenarios 1 and 2 were analyzed simultaneously.

Scenario 2: Dispersion Parameter

As discussed in Cook and Wei (2002), a larger dispersion parameter creates higher bias when estimating the mean rate for the *after* period (Λ_2), but it reduces the bias when we estimate the safety effectiveness. For a better understanding the extent of this bias, this scenario assumed that the dispersion parameters varied from 0.25 (small heterogeneity) to 7 (very large heterogeneity). For this scenario, the following variables were assumed to be fixed: a sample size equal to 100, and the safety effectiveness equal to 0.50. It should be pointed out that the dispersion parameters that have been observed with crash data rarely go beyond 2.0.

Scenario 3: Sample Size

Equation (2.31) was derived with the assumption that sample size tends towards infinite. There is a need to examine whether the estimator changes or not when sample size reduced. The scenario assumed that sample size varied as follows: 10 (small), 30 (medium), and 100 (large). For this scenario, the safety effectiveness was equal to 0.50.

Scenario 4: Safety Effectiveness

Based on equation (2.31), it is clear that larger values of the index of safety effectiveness cause higher site selection bias. The scenario assumed that safety effectiveness for three values: 0.90 (high), 0.70 (medium), and 0.50 (low). The sample size was equal to 100.

Scenario 5: The Standard Deviation of Safety Effectiveness

The effectiveness of treatment was assumed as a constant in Cook and Wei (2002), but it is not always true in practical. We assumed the variance of safety effectiveness may also be related to the site selection effect bias, and this scenario assumed a standard deviation of the index of safety effectiveness that varied from 0.05, 0.1, and 0.2. The safety effectiveness was made equal to 0.90, in order to avoid having the term θ equal to or less than zero. This scenario is more of a theoretical exercise, since the term θ is usually estimated from the data. This scenario could also be used to replicate the application of a crash reduction factor (CRF) characterized with different levels of uncertainty.

4.2.1.2 Simulation Protocol

The simulated data was generated using the software R (R Development Core Team, 2006)) .The general steps were as follows:

(1). For each dispersion parameter, generate the crash mean rate ($\hat{\lambda}_{i1}$), which follows a gamma distribution, $gamma \sim \left(\alpha^{-1}, \alpha\right)$, and generate a count with a Poisson mean ($N_{i1} \sim Poisson\left(\Lambda_{i1} = \hat{\lambda}_{i1} \times \tau_i\right)$) for each site i , where τ_i , the sample mean, was equal to 3. Sample mean values equal to 1, 5 and 10 were also tested, but the results are not presented here due to space constraints. All the results were consistent with the values presented in this dissertation, expect at the boundary when α is almost equal to zero.

(2). Generate three years of counts in the before period using Λ_{i1} for each site. Generate the data for *m* = 5,000 sites, but randomly select 100, 30 and 10 sites depending on the scenario.

(3). Only for Scenario 5: Generate the treatment effectiveness for each site using a normal distribution, $\theta \sim N\left(\mu, \sigma^2\right)$, where μ = 0.90 and σ^2 = 0.05, 0.10 or 0.20.

(4). Estimate the crash mean rate ($\hat{\lambda}_{i2}$) for each site in after period equal to the product of the above matrixes ($\hat{\lambda}_{i1} \times \theta$). Then, generate three years of count using the Poisson distribution, $N_{i2} \sim Poisson\left(\Lambda_{i2} = \hat{\lambda}_{i2}\right)$.

(5). Then, n sites are selected as the sample whose observed crash numbers are larger than the entry criteria (0,…, 5), and its effectiveness can be estimated using equations (3.5), (3.6), (3.7) or (3.8). The dispersion parameter can be estimated using equation (3.10) and (3.11).

(6). When the control group is used, θ is equal to 1. In other words, there is no difference in the mean rate between *before* and *after* periods.

(7). Repeat steps 2 to 6 for a total of 1,000 times, and estimate the various biases $\left(\theta - \hat{\theta}_{1000}\right)$.

It should be pointed out that the EB method was not used or evaluated by Cook and Wei (2002).

4.2.1.3 Scenario Results

This section describes the results based on the above simulation protocol. The results are presented for each scenario. As mention earlier, five possible factors include entry criteria, sample size, inverse dispersion parameters, safety-effectiveness values, and the standard deviation of safety effectiveness.

Scenario 1 − RESULTS

Figure 4-5 shows the site selection bias for the Naïve, CG, EB_{MM}, EB_{CG}, and the Adjusted methods. Overall, this figure shows that the bias goes down as the dispersion parameter increases, except when α is almost equal to zero (at least for C < 3) (recall that if $C = 0$, $N_{i1} \geq 1$, etc.). This was expected given the characteristics of equation (2.31). The greater the entry criteria the more biased the estimate will be. Among the four methods, the Naïve (Figure 4-5a) and the EB_{MM} (Figure 4-5c) methods are the ones that are the most affected by the site-selection bias; θ can be overestimated by as much as 36%. Based on equation (2.31), unless the dispersion parameter is close to infinite ($\alpha \rightarrow \infty$), θ will be biased if an entry criteria is used (i.e., the bias never equals zero when $\alpha = 7$). Readers may be surprised to see that the EB_{MM} method does not reduce or eliminate the RTM when site-selection effects are included. This is caused by the fact that the Method of Moment (MM) estimator is calculated using the characteristics of the truncated sample rather than the full population or non-truncated sample. Appendix A describes in greater detail the conditions under which the EB method (both for EB_{MM} and EB_{CG}) can be biased.

When a control group is used, the bias might be theoretically eliminated. For the CG method (Figure 4-5(b)), the control group needs to have the same characteristics (i.e., the same sample mean and variance (which can be used for obtaining the dispersion parameter) as the truncated sample used for the Naïve method (see right-hand side of Figure 2-3). As explained above, it may be difficult in practice to find datasets with the exact same characteristics as those for the treatment group. For the EB_{CG} method (Figure 4-5d), the control group needs to have the same characteristics as those of the full sample (or sample population) from which the truncated data were used for calculating the Naïve or EB_{MM} estimates (see left-hand side of Figure 2-3). Again, the reader is referred to Appendix A for the conditions under which the EB_{CG} can be biased.

The application of the control group is further explored in Figure4-6. Figure4-6(a) is the same as Figure 4-5(b) and is used to compare the results with the other figure below. Figure4-6(b) shows that when the control group does not have the same characteristics – in this case the same sample mean – the site-selection bias is still present (although, it is still smaller than when using the Naïve before-after method). Furthermore, the bias is also in the opposite direction (underestimate θ).

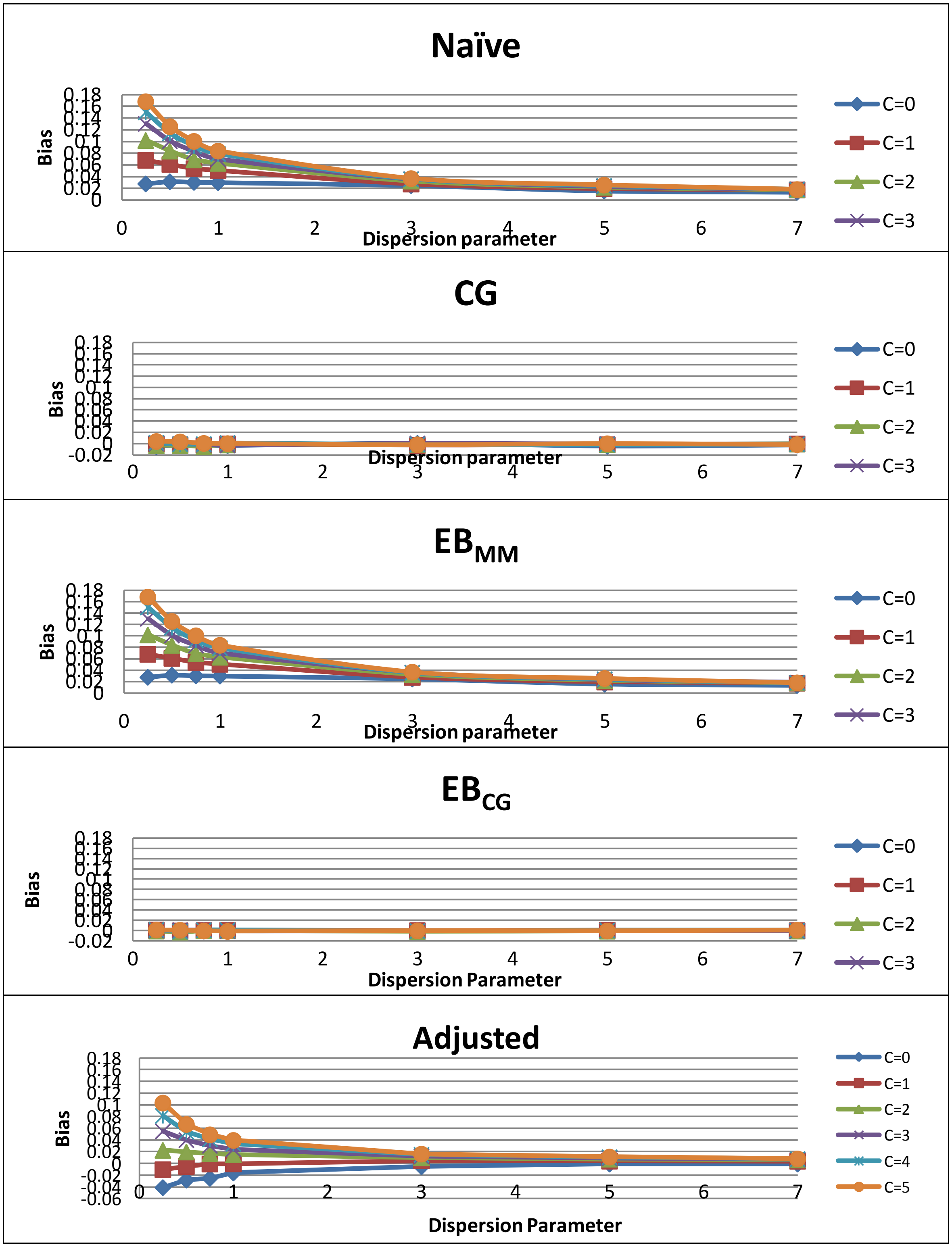


Figure 4-5 Site selection bias for the Naïve, CG, EB and Adjusted methods

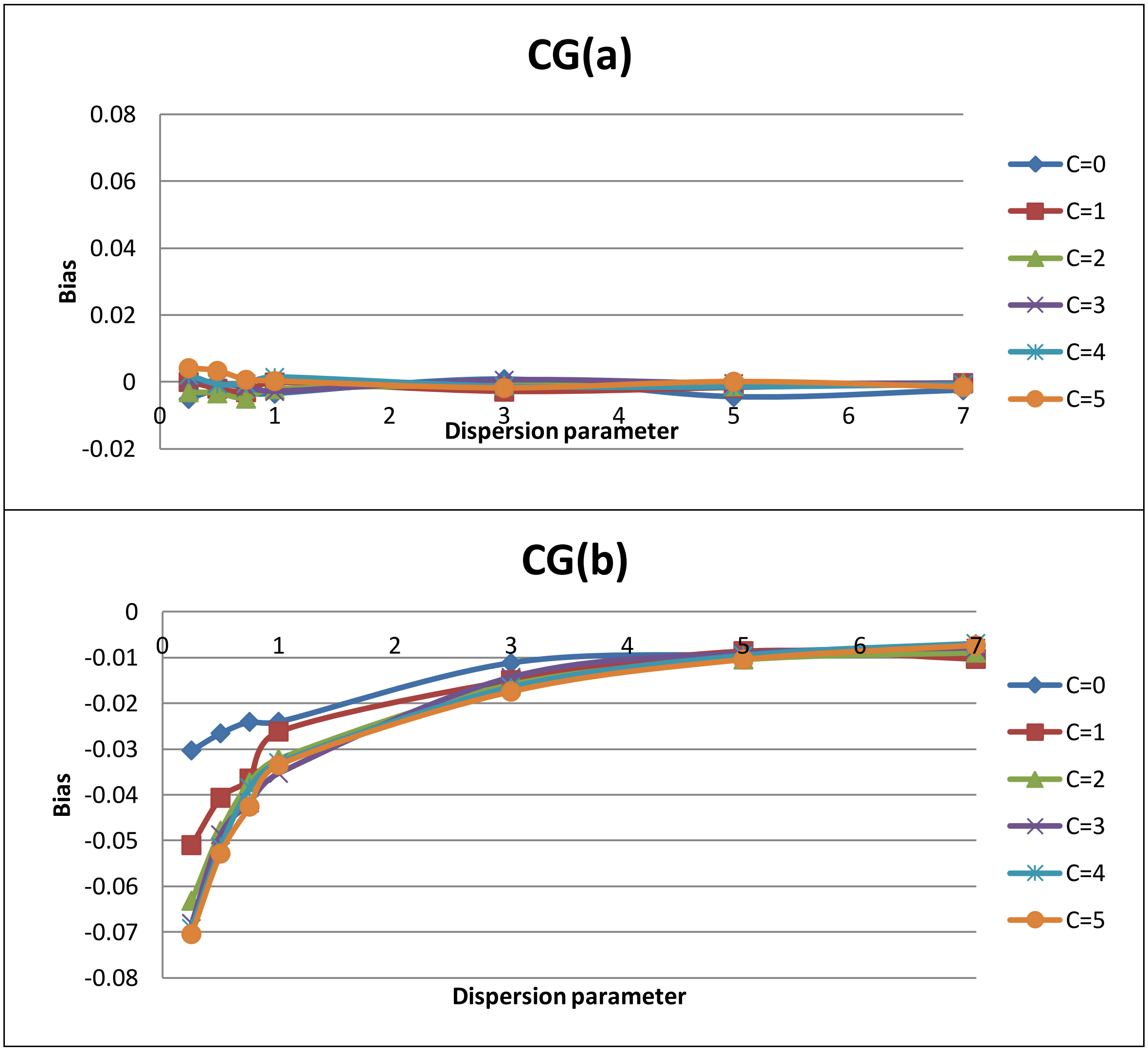


Figure4-6 Site selection bias for the CG method for the following characteristics: (a) same mean, (b) 0.75 × mean

Scenario 2 − RESULTS

As discussed in the first scenario, the site selection bias decreases when the dispersion parameter increases, but the rate at which the bias decreases becomes almost flat for values above 5 (Figure 4-6). With the Naïve method, the bias is never eliminated, compared to the CG and EB_{CG} method. It should also be noted that when the dispersion parameter tends towards zero (which now almost becomes a Poisson model), the site selection bias still exists, as pointed out by Cook and Wei (2002).

Scenario 3 – RESULTS

Figure 4-7 shows that the sample size related to the treatment group does not affect the bias considerably. When the sample size is over 30, the bias estimated for the Naïve method, using equation (3.5), and the one estimated with equation (2.31) are very close. Furthermore, there is a slight difference between the simulated bias and the estimated one when the dispersion parameter is less than 1 (most often observed in crash data). However, the maximum difference (sample size=10, C=0) is about 0.04 when the safety effectiveness is equal to 0.50. Hence, equation (2.31) which was derived by assuming that the sample size is close to infinity (∞) may be used for estimating site selection biases even the sample size is small.

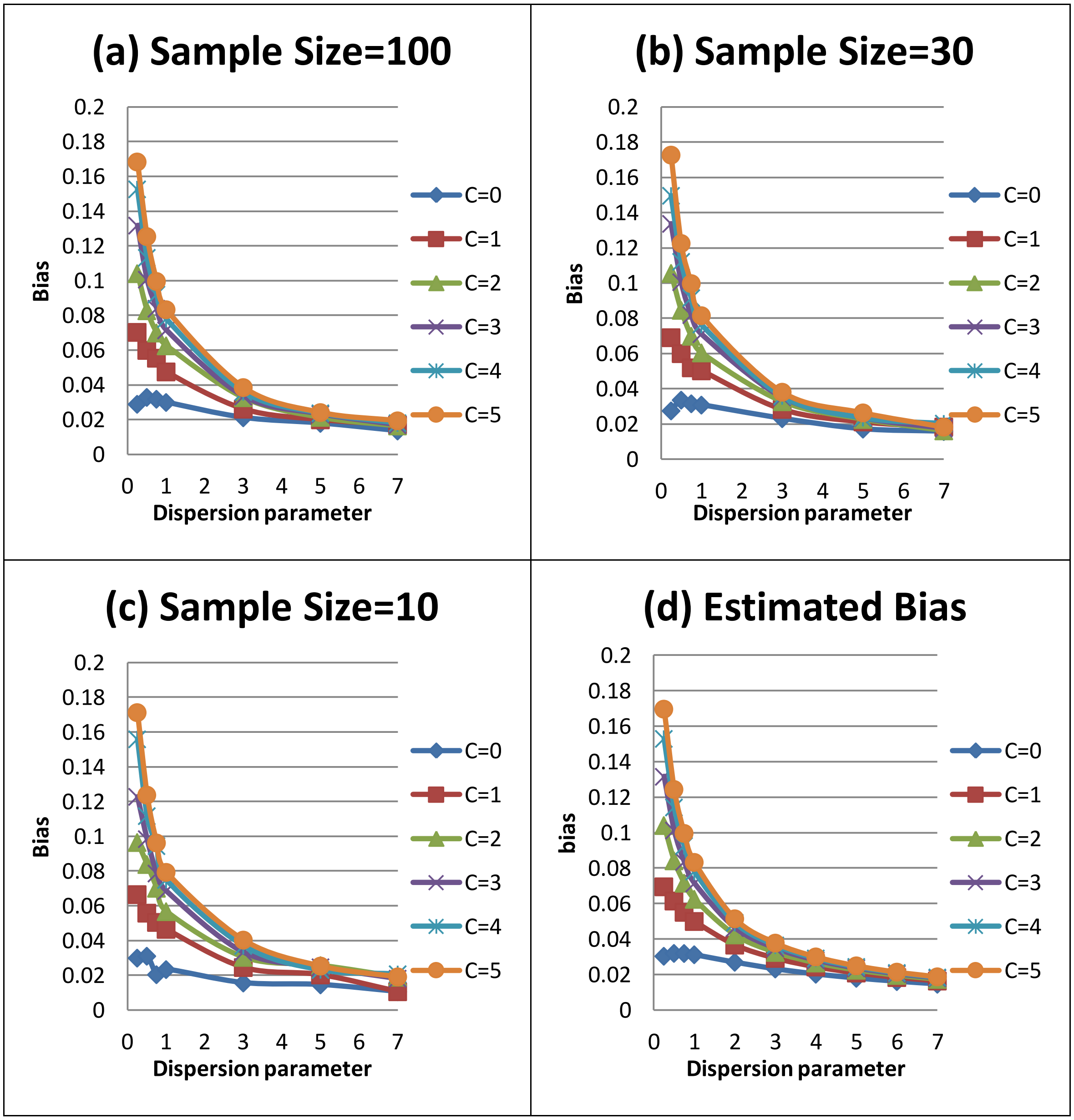


Figure 4-7 Site selection bias for the Naïve method when the sample size is equal to (a) 100, (b) 30, (c) 10 or (d) estimated using equation (2.31).

Scenario 4 – RESULTS

Figure4-8 shows that the value of the index of effectiveness (θ) changes the value of the site-selection bias, because the bias increases when the countermeasure effectiveness becomes smaller (closer to 1). This figure also shows that the ratio between the site-selection bias and the safety effectiveness seems fixed, which is consistent with the characteristics of equation (2.31). The results illustrate that the selection bias may influence the final decision made by the engineer or transportation safety specialist when different treatments are evaluated, but only a single or a limited number of treatments can be selected. For instance, when the effectiveness of two treatments is very close but when their entry criteria are very different (because different sites are used), the selection bias may overestimate the performance for one of the treatments over the others, which could result in the wrong selection for the treatment. Looking at the first line (the orange line with the dashed circled node) in Figure4-8 (a), and the last line (blue line with the dashed circled node) in Figure4-8 (b), the biased estimator in Figure4-8 (a) is equal to 0.6 (=0.90-0.30), which is lower than the estimator (0.65=0.70-0.05) found in Figure4-8 (b). This means that the treatment identified in Figure4-8 (a) would be selected over the treatment identified in Figure4-8 (b), since the former one looks like it reduces more crashes, although it in fact reduces fewer crashes than the one shown in Figure4-8 (b).

Scenario 5 − RESULTS

Figure 4-9 shows that the standard deviation associated with the countermeasure effectiveness does not change the value of the bias significantly. However, if the standard deviation is relatively large, such as when it is equal to 0.20, the simulated results become unreliable, because the mean value for the after period may be equal to or less than zero. When this occurred, the observations were removed from the analysis.

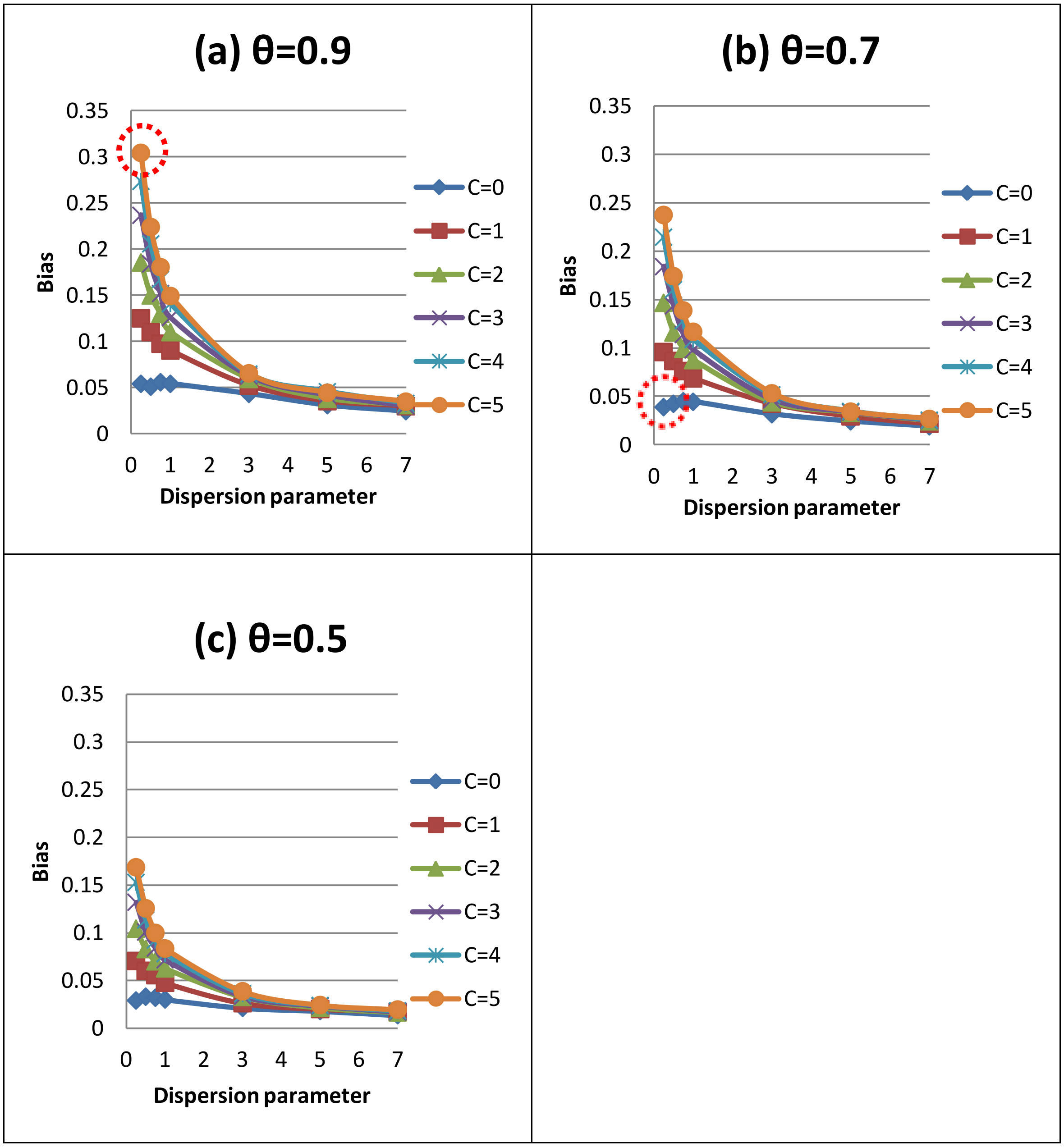


Figure4-8 Site selection bias for the Naïve method when the safety countermeasure is equal to (a) 0.90, (b) 0.70 and (c) 0.50

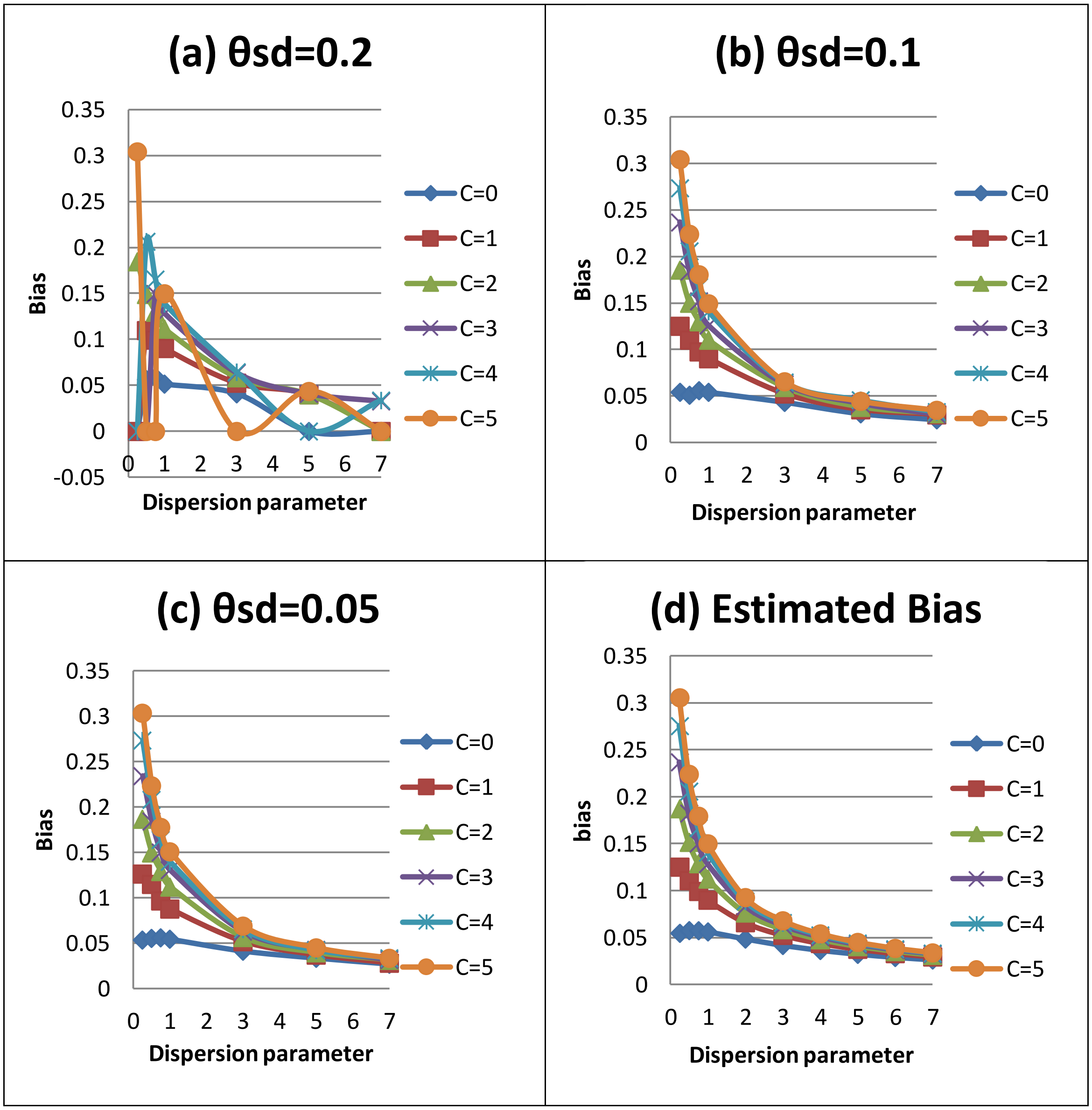


Figure 4-9 Site selection bias for the Naïve method when the standard deviation is equal to (a) 0.2, (b) 0.1, (c) 0.05 or (d) the bias is estimated using equation (2.31).

4.2.2 Difference

Aside from the index of safety effectiveness (θ), difference (δ) is another possible index that can be used for estimating an intervention's effectiveness for count data. This section shows the simulation result for estimating difference using various before-after methods and possible influence factors. Simulation protocols are not repeated here since they are same as Section 4.2.1.2.

4.2.2.1 Scenarios for Possible Factors

In the following subsections four possible factors of influence will be discussed by sensitive analysis: entry criteria, different before-after methods, inverse dispersion parameters, and difference value. We did not repeat scenarios for various sample sizes or the standard deviation of difference, since the previous section already shows that they are unrelated factors.

As mentioned above, note that the estimators were estimated in the same way as for the subsequent scenarios. For these scenarios, the following variables were assumed to be fixed: a sample size equal to 100 and a difference equal to 1.5. Note that Scenarios 1 and 2 were analyzed simultaneously.

Scenario 1: Direct Comparison of the Methods

The setting is similar to section 4.2.1. The reason why we repeat it here is because equation (2.34) shows that using CG method (with ideal control group data) might not be able to remove site-selection bias on difference, and it is different from the results for safety effectiveness.

Scenario 2: Dispersion Parameter

Same as Section 4.2.1, the reason that we repeat it again here is because equation (3.15) shows that a higher dispersion parameter causes higher site-selection bias on the difference (δ), while a higher dispersion parameter causes lower site-selection bias on the index of safety effectiveness (θ).

Scenario 3: Difference

Based on equation (2.30), it is clear that larger values of the index of difference cause higher site-selection bias. Also, site-selection bias is zero when difference is zero, which indicates that the treatment did not work. The scenario assumed that difference for three values: -2.1 (high), -1.5 (medium), and -0.9 (low).

4.2.2.2 Scenario Results

This section describes the results based on the above simulation protocol. The results are presented for each scenario. As mention earlier, four possible factors of influence will be discussed by sensitive analysis: entry criteria, different before-after methods, inverse dispersion parameters, and difference value.

Scenario 1 – RESULTS

Figure 4-10 shows the site-selection bias for the Naïve, CG, EB_{MM}, EB_{CG}, and the Adjusted methods. Overall, this figure shows that the bias increases as the dispersion parameter increases. This was expected given the characteristics of equation (2.30). The greater the entry criteria the more biased the estimate will be. Among the five methods, the Naïve (Figure 4-10 (a)) and the EB_{MM} (Figure 4-10 (c)) methods are the ones most affected by the site-selection bias; δ can be overestimated by as much as 533%. As discussed by Cook and Wei (2002), δ will be biased when an entry criterion is used. Readers may be surprised to see that the CG and EB_{CG} remove just a few site-selection biases on the estimator of difference, while these two methods remove all site-selection bias on the estimator of safety effectiveness. This simulation result is consistent with equation (2.34), because the site-selection bias generated by using the CG method is equal to $\frac{\delta(\mu_1\alpha+1)}{(\Lambda_1\alpha+1)}$. The application of the control group is further explored in Figure 4-11. Figure 4-11(a) is the same as Figure 4-10 (b), and is used to compare the results with the other figure below. Figure 4-11(b) shows that when the control group has a lower mean (0.75 × mean), the site-selection bias is slightly smaller than the bias generated from the Naïve method. Conversely, using a control group that has a higher mean value (1.2 × mean) causes a site-selection bias that is even higher than from using the Naïve method (Figure 4-11(c)). In other words, the site-selection bias caused by using a dissimilar control group might be even higher than with using just the Naïve method.

Also, in the same way as the Naïve method, the EB_{MM} method does not reduce any site-selection effects, because they used the characteristics of the truncated sample rather than the full population or non-truncated sample. Appendix B describes in greater detail the conditions under which the CG and EB methods (EB_{MM} and EB_{CG}) can be biased.

Finally, using the Adjusted Method reduced site-selection bias by about 40%—even when the biased estimators of α, and Λ are used, or when they are not known with certainty. Compared to each other method, the Adjusted Method provides a more precise estimate than the Naïve method and performs better than the CG and ANCOVA methods when similar control group data are not available. It should be pointed out that the Adjusted Method will not completely eliminate the site-selection effects unless α, and Λ are fully known. Obviously, these values are rarely known in practice.

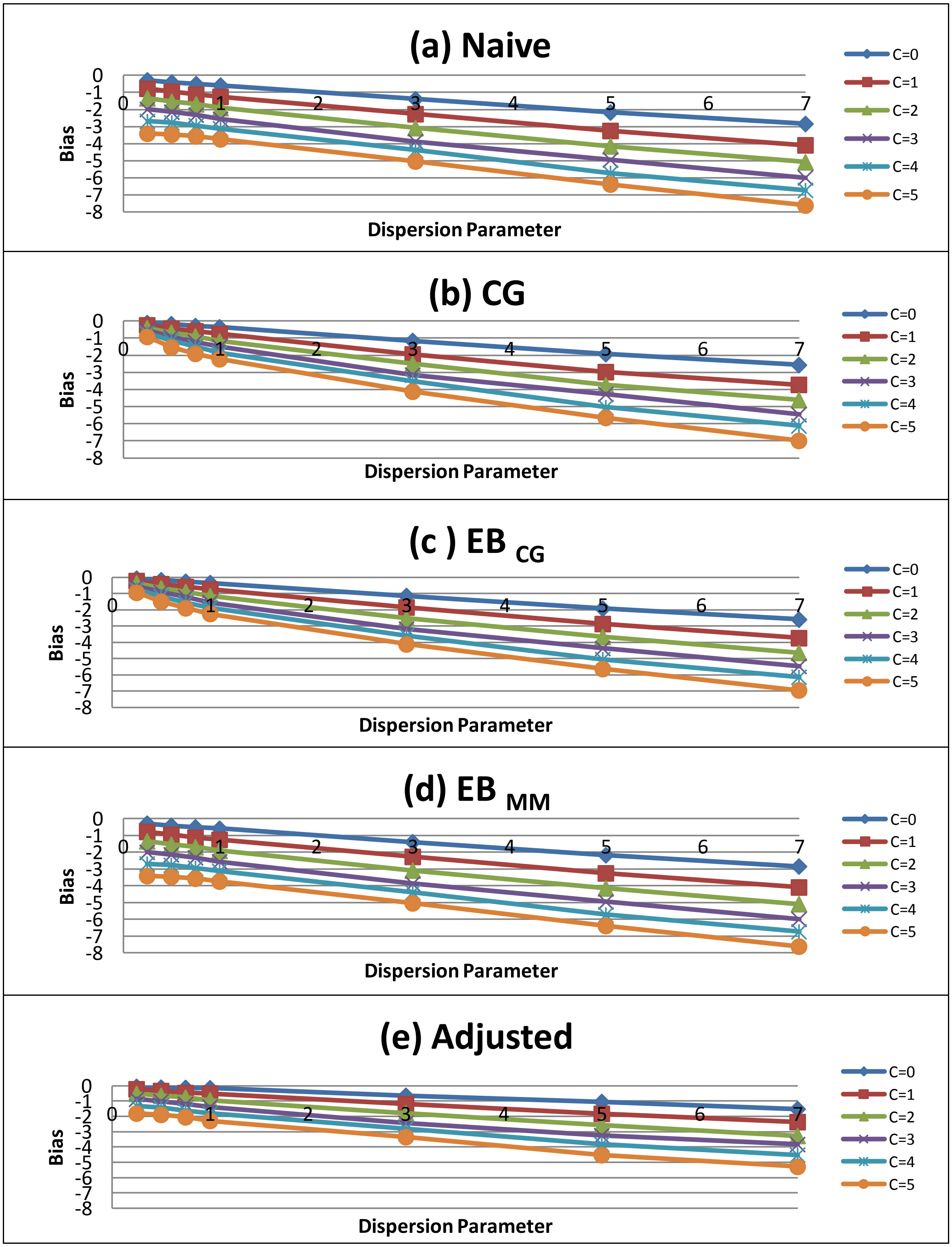


Figure 4-10 Site selection bias for the Naïve, CG, EB and Adjusted methods

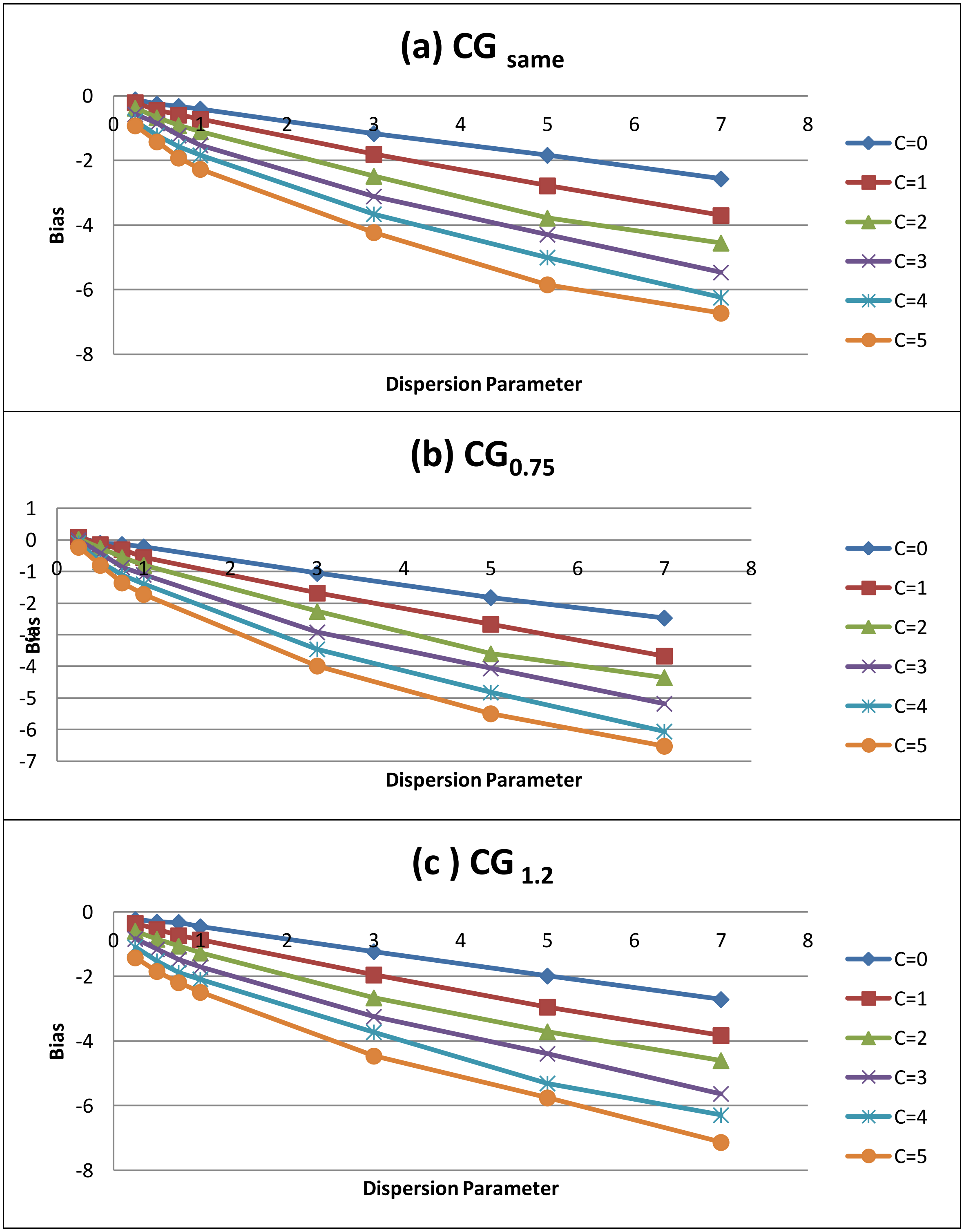


Figure 4-11 Site selection bias for the CG method for the following characteristics: (a) same mean, (b) 0.75 × mean, and (c) 1.2 × mean

Scenario 2 − RESULTS

As discussed in the first scenario, the site-selection bias increases when the dispersion parameter increases, and the rate at which the bias increases becomes sharper for values above three (Figure 4-10). With the Naïve method, the bias is never eliminated even when dispersion parameter is close zero, as compared to the CG method.

Scenario 3 – RESULTS

Figure 4-12 shows that the value of the difference, (δ), changes the value of the bias, because the bias increases when the magnitude of difference increases (i.e. becomes more negative). This figure also shows that the ratio between the site-selection bias and the difference seems fixed, which is consistent with the characteristics of equation (3.15). The results illustrate that the selection bias may influence the final decision made by the engineer or transportation safety specialist when different treatments are evaluated; however, only a single or a limited number of treatments can be selected. For example, when the difference between two treatments is very close but when their entry criteria are very different (because different sites are used), the selection bias may overestimate the performance for one of the treatments over the others, which could result in the wrong selection for the treatment. Looking at the first line (the orange line with the dashed circled node) in Figure 4-12 (a), and at the last line (the blue line with the dashed circled node) in Figure 4-12 (e), the biased estimator in Figure 4-12 (a) is equal to -5.9 (=-0.9-5), which is lower than the estimator (-5.1=-2.1-3) found in Figure 4-12 (e). This means that the treatment identified in Figure 4-12 (a) would be selected over the treatment identified in Figure 4-12 (e), since the former one looks like it reduces more crashes, although it in fact reduces fewer crashes than the one shown in Figure 4-12 (e). However, using our adjusted method may solve this problem by removing partial selection bias. The biased estimator in Figure 4-12 (b) is equal to -4(=-0.9-3.1), which is close to the estimator (-4=-2.1-1.9) found in Figure 4-12 (f).

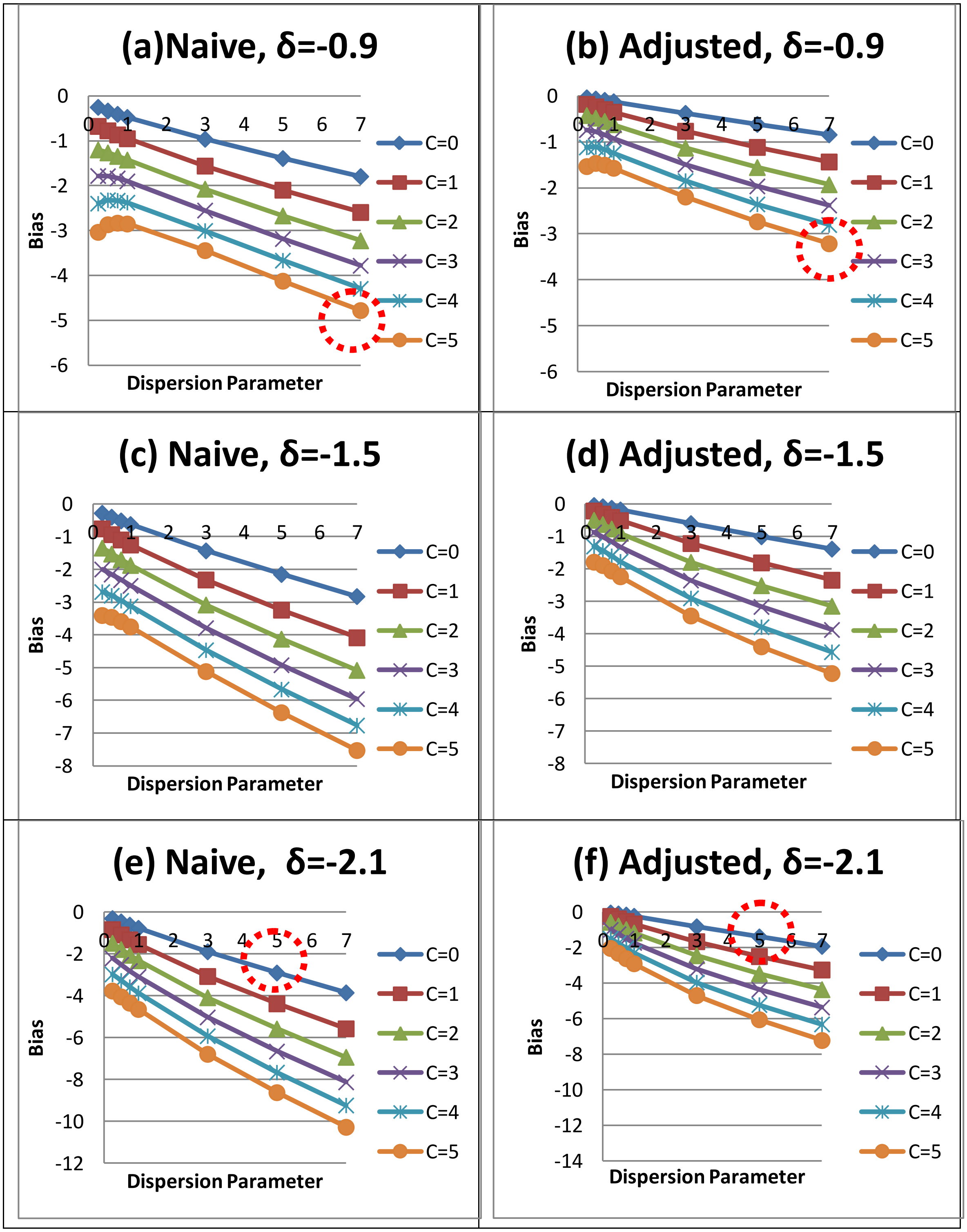


Figure 4-12 Site selection bias for the Naïve and Adjusted method when the difference is equal to (a) -0.9, (b) -1.5 and (c) -2.1

4.2.3 Dispersion Parameter

As discussed above, the dispersion parameter can be influenced by the site-selection effects. This section is used to show how simulating site-selection bias affects the estimator of dispersion parameter. Based on equations (2.37) and (3.11), the possible related factors are entry criteria, dispersion parameter, and estimation methods. Simulation protocols are not repeated because they are same as Section 4.2.1.2. According to the simulation result, Figure 4-13 shows that higher entry criteria and a higher dispersion parameter may cause higher selection bias, which results in an underestimated dispersion parameter. However, when the true dispersion parameter is unknown there are no significant differences between our adjusted estimators and the naïve estimator. In other words, our site-selection bias estimation, equation (3.11), cannot remove a major portion of selection bias because of using naive dispersion parameter instead of the true dispersion parameter. Our adjusted method only works when dispersion parameter is known. Further studies are necessary to solve this problem.

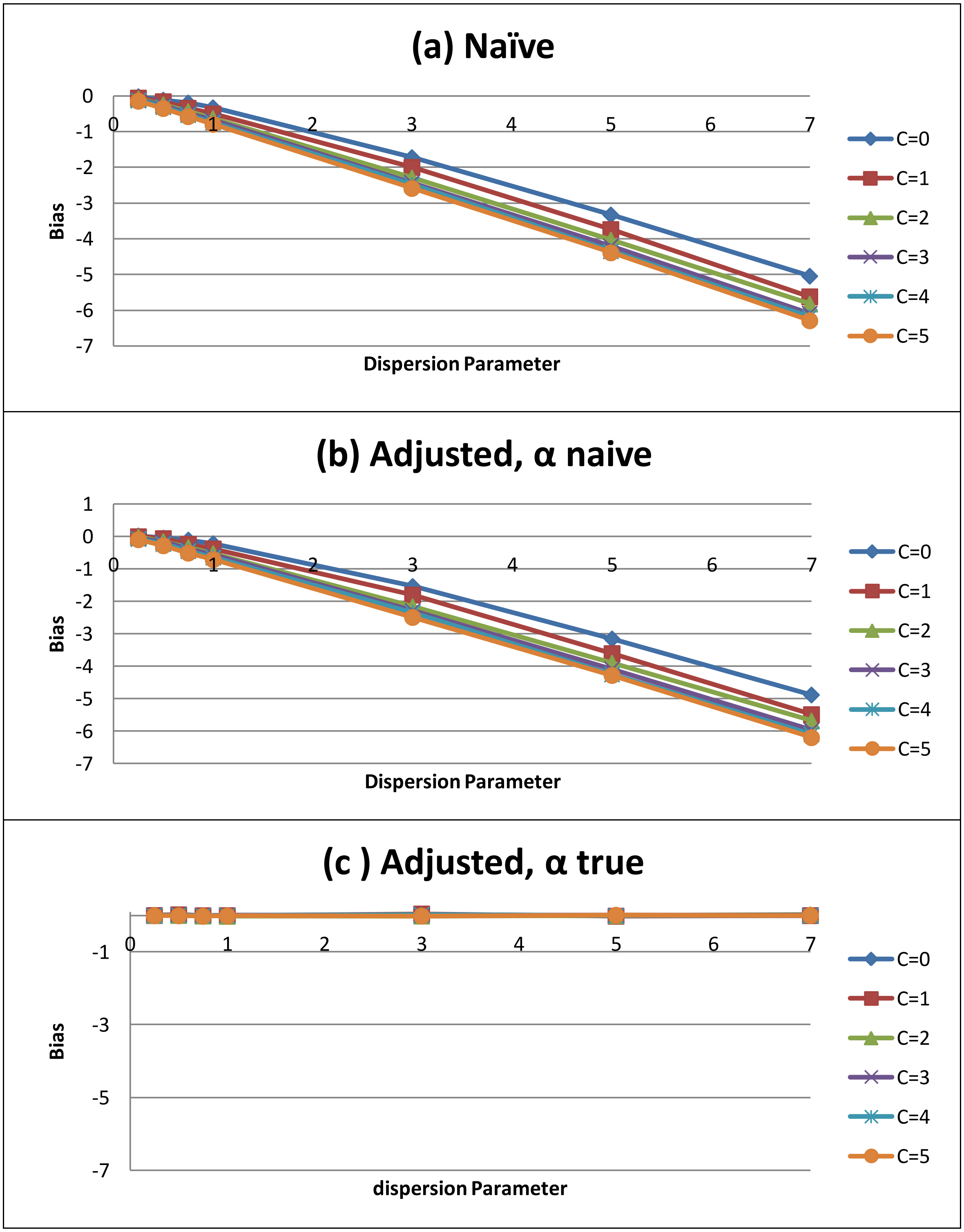


Figure 4-13 Site-selection bias for the (a) Naïve, (b) Adjusted method using naïve α and (c) using true α

4.3 Count Data: Simulation from Varying Crash Mean

In the previous section (Section 4-2), the simulation results showed that equation (3.8) would correctly estimate site-selection bias when the crash rate was assumed to be a fixed value (i.e., when all the sites had the same mean) and the crash counts follow a Poisson-Gamma model. However, this assumption regarding a fixed mean might not be true in practice, especially when sites are far from each other or have different characteristics, such as in their number of lanes, intersection types, or traffic volumes. Hence, this section changes the equation such that the crash rate is not constant, and instead varies from one site to the next. Two different settings will be examined below: (1) crash rates that follow a log-normal distribution and, (2) crash rates that are based on the output of a regression model.

4.3.1 Crash Rates Followed a Log-normal Distribution

The above-mentioned scenario assumed that the number of crashes followed a Poisson Gamma-Lognormal distribution, and the setting was kept the same as that which was used in Lord (2006). The variance (σ) of the Lognormal distribution was varied as follows: 0.01 (small heterogeneity), 0.5 (median heterogeneity), and 1 (very large heterogeneity). The other input variables, such as mean value and dispersion parameter, were kept the same as those described in Section 4.2.1.2. The simulation protocol was as follows:

$$\begin{aligned} &N_{ik} \sim Poisson(u_i \Lambda_k) \\ &\Lambda_1 = \lambda_1 \times t_1 \\ &\lambda_1 \sim \log normal(\log(3), \sigma) \\ &u_i \sim Gamma(\alpha^{-1}, \alpha) \end{aligned} \tag{4.1}$$

Figure 4-14 shows the simulation results for different variances.

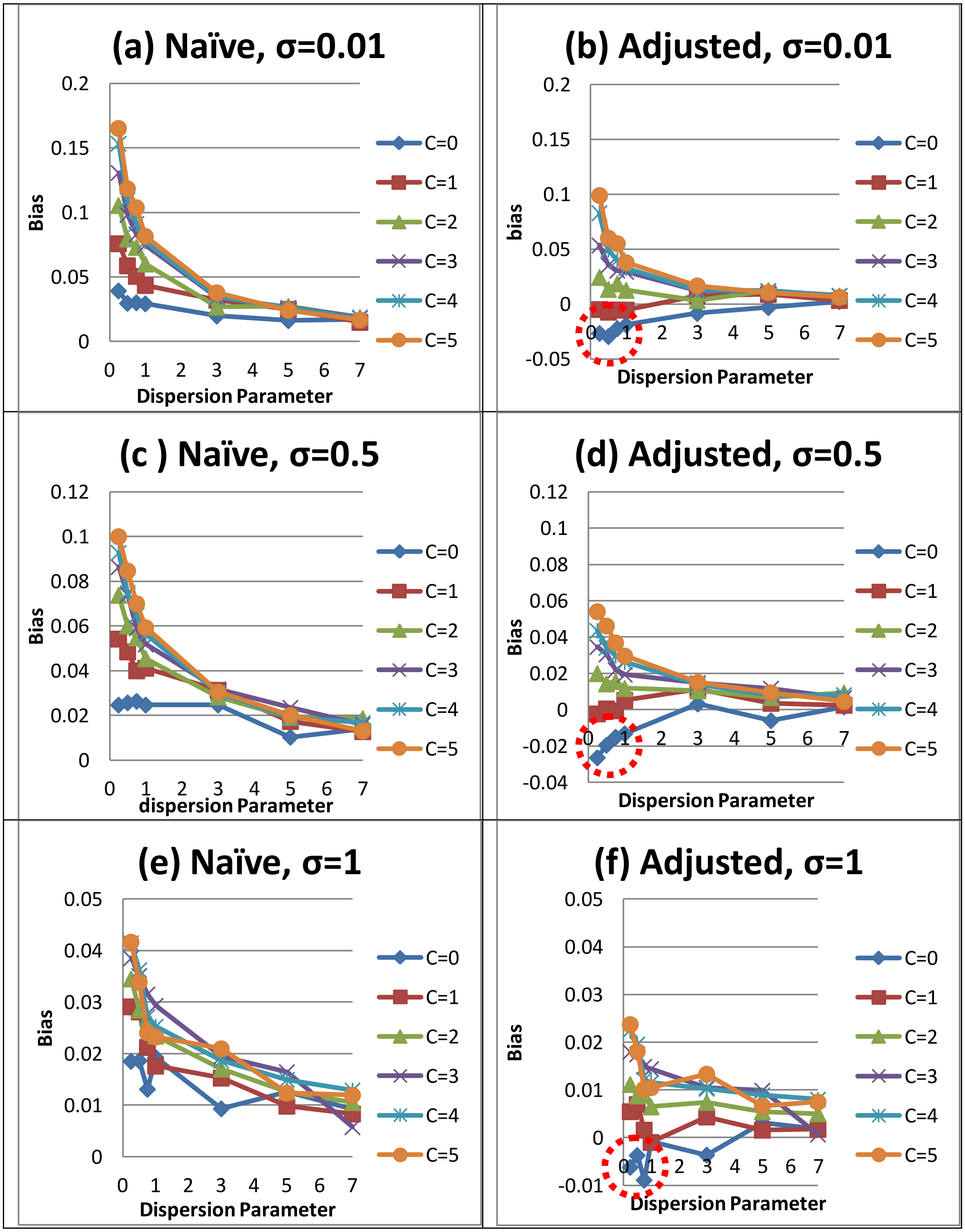


Figure 4-14 Site-selection bias for Naïve and Adjusted methods when the standard deviation of the crash mean is equal to 0.01, 0.5, and 1

Figures 4-14 clearly shows that larger variances in the mean distributions decrease site-selection biases. This result was expected because larger variances among the crashes mean increase in the between-subject variance (similar to the effects of larger dispersion parameter values), resulting in lower site-selection biases. However, it should be pointed out that the adjusted method, demonstrated in equation (3.8), still reduced the site-selection bias by approximately 50%. However, the estimator might overestimate the site-selection bias when the entry criteria are low.

4.3.2 Crash Data Based on a Regression Model

In this section, the crash regression model proposed by Harwood et al.(2007) was used to generate the crash dataset. The simulation protocol was as follows:

$$\begin{aligned} &Y_i \sim Poisson(u_i) \\ &u_i = e^{\beta_0}(F_1)^{\beta_1}(F_2)^{\beta_2} \end{aligned} \tag{4.2}$$

where

β_0 : intersect is -10.63

β_1: major road's parameter is 1.07

β_2 : minor road's parameter is 0.23

F_1: traffic volume of major roads which follows N (9986, σ_{crm})

F_2 : traffic volume of minor roads which follows N (3474, σ_{crm})

In Harwood et al. (2007), the traffic volumes averaged from Minnesota and North Carolina were 21,033 and 7,317.5 vehicles per day for major and minor roads, respectively. In order to get the same overall mean crash rate (=3) as was achieved in the previous section, the researcher adjusted the mean of traffic volume to 9,986 and 3,474 vehicles per day, both followed a normal distribution with a variance, σ^2_{crm} . The standard deviations (σ_{crm}) of the flow were set as follows: 50 (small), 500 (median), and 3000 (very large). Figure 4-15 shows the simulation results for the different variances. Generally, a larger variance associated with the mean crash rate caused a lower site-selection bias. In sum, site-selection biases were lower when the crash mean was not fixed but the reductions were not obvious until the variance associated with the mean became extremely large. This conclusion will have a significant impact for safety analyses, since the sample sites of the treatment or the control groups do not have the same means, but their variances of mean are not extremely large. In other words, because researchers usually exercise control over all the input variables in order to keep their samples homogeneous , their estimates may still be biased.

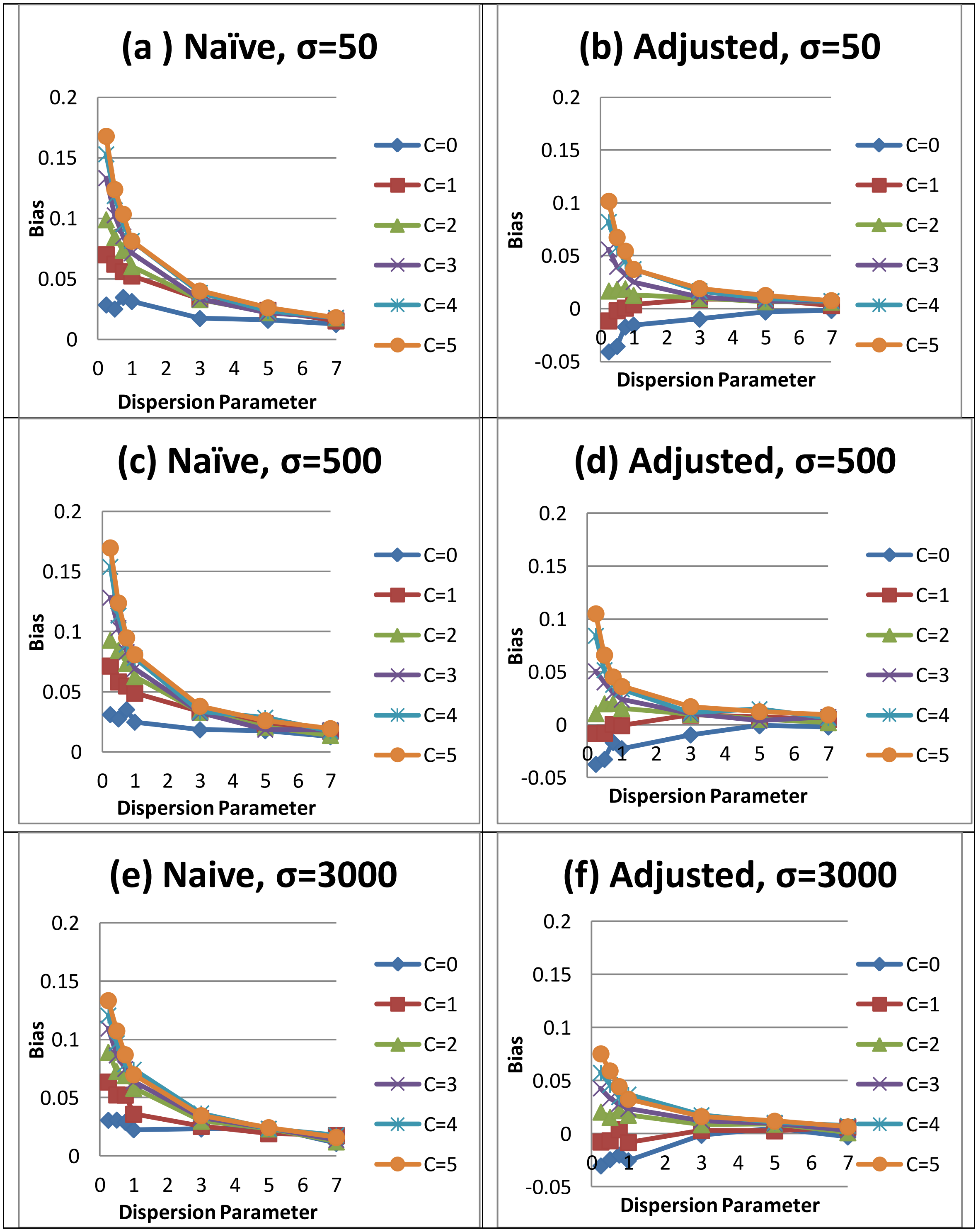


Figure 4-15 Site-selection bias for the adjusted method when the standard deviation of traffic volume in crash regression model is equal to (a) 50, (b) 500, and (c) 3000

4.4 Summary

This section has successfully used simulation data to show how setting an entry criterion influences the estimation of traffic-safety countermeasures and dispersion parameter for continuous data (δ) and count data (δ, θ, α), respectively. For continuous data, four scenarios were evaluated, directly comparing the methods, differences in the sample sizes, within-subject variances, and between-subject variances. For the count data, five scenarios were evaluated, directly comparing the methods, differences in the sample sizes, differences in the dispersion-parameter values, different safety-effectiveness values, and difference in the standard-deviation values associated with safety effectiveness. To test and evaluate how each of the biases would work in practice, crash data with a varying mean were also simulated.

Based on the simulated scenarios, the study results showed that among all methods evaluated, the Naïve and EB_{MM} (only for count data) methods were most significantly affected by the selection bias. Using other advanced methods (such as the control group (CG), EB_{CG}, or ANCOVA method) might eliminate the site-selection bias on difference (for continuous data) and safety effectiveness (for count data), as long as the characteristics of the control group are exactly the same as those of the treatment group. In practice, however, this might not be possible. Also, the simulation results illustrated the conditions required for causing higher site-selection biased for continuous data (higher entry criteria, higher within-subject variance, and smaller between-subject variance) and for count data (higher entry criteria, larger values of the index (θ), and smaller dispersion parameter values). The above findings are consistent with bias estimation equations (2.8) and (2.45). For count data, using the CG or EB_{CG} method did not eliminate the site-selection bias with regard to estimates of difference, even with a perfect control group, and the estimates were very close to Naïve and EB_{MM} estimates. Moreover, the higher dispersion parameter caused a higher site-selection bias on difference while it caused lower selection-bias on safety effectiveness, because increasing the value of dispersion parameter raised $\mu_1(=\hat{\Lambda}_1)$ more than $\mu_2(=\hat{\Lambda}_2)$.

Finally, with regard to dispersion parameter, setting higher entry criteria and higher dispersion parameter could cause a higher site-selection bias, which would result in an underestimation of dispersion parameter. The results have shown that the Adjusted method, as illustrated in equations (3.8) and (3.15), could partially eliminate site-selection biases in estimations regarding safety effectiveness and difference, even when biased estimators of the mean and dispersion parameter of a truncated Negative Binomial distribution are used, or when the crash rate is not a fixed number. However, the Adjusted method would not work well to reduce the site-selection bias for estimates of the dispersion parameters. Thus, further study is necessary to solve this problem. To verify the simulation results, the following sections document the use of observed data to examine the site-selection biases for continuous data and count data, respectively.

5. APPLICATION USING OBSERVED DATA

In order to better illustrate the simulation results, the bias-estimation equations (3.4), (3.8), (3.11), and (3.16) were applied to two different observed datasets: continuous data (Section 5.1) and count data (Section 5.2). For continuous data, equation (3.4) was used to assess the effects of increasing the speed limit on multilane highways in Florida, since normal distribution is one of most common distributions used to characterize speed data. The dataset was first collected and explored by Muchuruza and Mussa (2004), whose findings did not completely eliminate the site-selection bias. The same dataset subsequently was analyzed by Park and Lord (2010) to examine the effects of the RTM in the before-after evaluation of the continuous data. For count data, two datasets were used, both of which followed a negative binomial distribution. The first dataset was collected in College Station, Texas. A dummy variable was used to simulate a hypothetical treatment that was implemented using the crash data collected in that city. The second dataset was assembled to evaluate the effects of cameras used to record red light violations on the overall number of crashes. The above datasets were used to evaluate the accuracy of the Adjusted method for estimating site-selection biases for various types of data that have different mean and sample-size values.

5.1 Continuous Data Collected in Florida

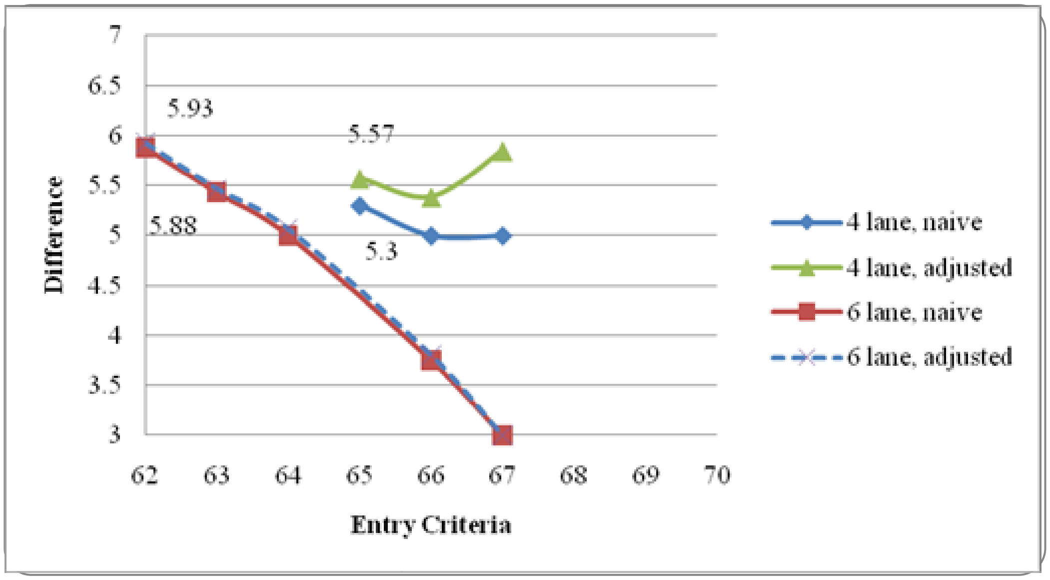
This section describes the application of the bias-estimating equations to observed speed data collected in Florida. Muchuruza and Mussa (2004) examined the impacts on observed driving speeds when the speed limit was increased from 65 mph to 70 mph on four- and six-lane highways. Their analysis was carried out at 18 different study sites, including 10 four-lane and 8 six-lane highways. Table 5-1 summarizes the important characteristics of this data. Using the Naïve before-after method, these researchers showed that the average speed increased by 5 mph. To examine the effects of the site-selection bias, equation (3.4) was used to obtain an adjusted difference. The first entry criterion was assumed to be 62, because the suggested initial assumption for the entry criteria should be equal to the smallest observed data (i.e., C=min Y_{ij}) (Johnson et al., 1970). Then, minimum entry criteria that were equal to 63, 64, 66, 67, and 68 were employed. Four-lane and six-lane highways were initially analyzed separately and were then combined to increase the sample size. By grouping the data, the adjusted estimators for six-lane highways could be compared with the results described in Park and Lord (2010).

Table 5-1 Before-After Speed Data (Muchuruza and Mussa, 2004)

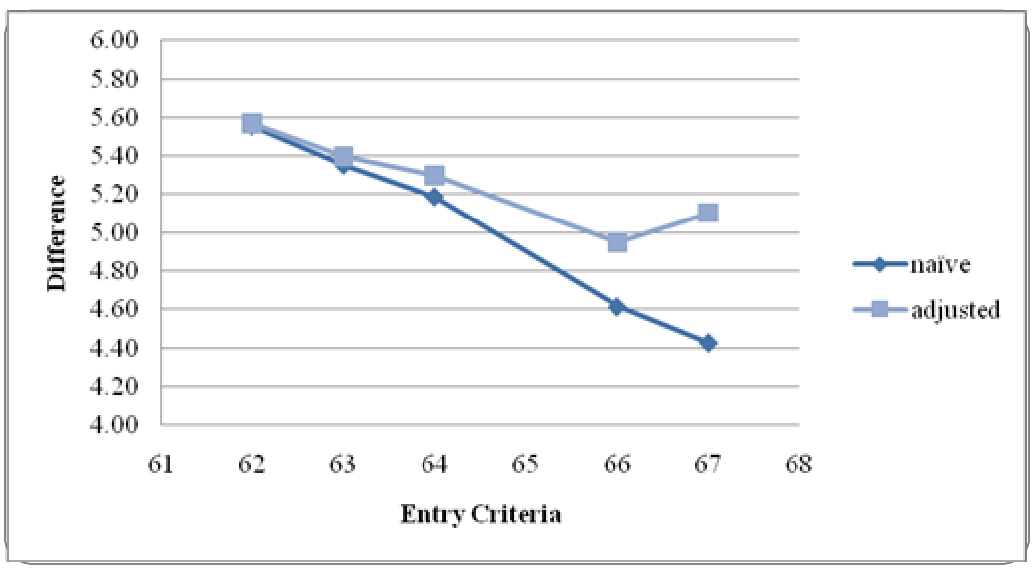
	Highway	Before (1996)	After (2002)	Before speed (mph)	After speed (mph)	difference
		Location, direction				
Four Lane Highway	I-75	At mile marker 89, WB,	Site 351,WB	66	74	8
		At mile marker 89, EB,	Site 351, EB	68	78	10
	I-10	Overpass E. of SR 85,WB,	Site 9901, WB	67	74	7
		Overpass E. of SR 85, EB,	Site 9901, EB	69	74	5
		C-280 overpass, WB,	Site 9901, WB	68	74	6
		C-280 overpass, EB,	Site 9901, EB	67	74	7
		Between SR257 & US221, WB,	Site 9928, WB	67	70	3
		Between SR257 & US221, EB,	Site 9928, EB	69	71	2
		East end of Aucilla River, WB,	Site 9928, WB	67	70	3
		East end of Aucilla River, EB,	Site 9928, EB	69	71	2
	Mean			67.7	73	5.3
Six Lane Highway	I-75	Between I-10 & CR136, NB,	Site 320, NB	66	73	7
		Between I-10 & CR136, SB,	Site 320, SB	66	74	8
		Between CR234 & SR21, NB,	Site 9904, NB	68	71	3
		Between CR234 & SR21, SB,	Site 9904, SB	67	71	4
	I-95	Between CR210 and I-295, NB,	Site 9905, NB	67	72	5
		Midpoint CR210 and I-295, SB,	Site 9905, SB	63	72	9
		Near Flagler CL, NB,	Site 9905, NB	69	72	3
		Near Flagler CL, SB,	Site 9905, SB	64	72	8
	mean			66.25	72.125	5.88

Figure 5-1(a) shows the results from the Naïve and Adjusted estimators for both four-lane and six-lane highways. This figure clearly illustrates that the estimators for four-lane highways have a higher selection bias than those for six-lane highways. For example, when the entry criterion is equal to 62, the Adjusted estimator (δ=5.93, removing site-selection bias) for the six-lane group is just slightly higher than the Naïve estimator (δ=5.88). However, when the entry criterion is equal to 65, the Adjusted estimator (δ=5.57) from the four-lane group is significantly higher than the Naïve estimator (δ=5.30). For an entry criterion equal to 66, the site-selection bias for the four-lane road group is 0.38 (δ=5.38-5.0) while the site-selection bias for the six-lane group is 0.05. Although the absolute difference appears to be small, the relative difference in percentages is actually large enough (7.6%) that the bias needs to be accounted for; Lord (2006) includes additional details about the impact of a biased estimate even when the absolute difference is small. Furthermore, the true site-selection bias could be as much as 15.2% (=7.6% $\times 2$) or more, since the adjusted method only captures 50% of the site-selection bias when biased estimators of mean and variance are used.

When the entire sample is analyzed, the difference between the Naïve estimators and adjusted estimators becomes more obvious (Figure 5-1 (b)). The Adjusted method increases the Naïve estimator by partially removing the selection bias, while the average adjusted estimator from the Park and Lord (2010) study remains the same (note that in Park and Lord's research the effects linked to the RTM are removed from each *site's* estimator, and not from the average estimator). Moreover, Figure 5-1 shows that higher entry criteria result in cause higher selection biases, a condition which is likely to exacerbate the tendency to underestimate the differences even more. Equation (3.4) partially eliminates selection bias but it cannot remove it all (for the reason discussed above). It should be noted that differences are underestimated when the true difference is positive ($|\delta_{naive}| < |\delta_{adjusted}|, \delta > 0$). On the other hand, differences are overestimated when the true difference is negative ($|\delta_{naive}| > |\delta_{adjusted}|, \delta < 0$). Generally, Figure 5-1 (a) and (b) support the theoretical points and simulation results described above: higher criteria cause larger (more negative) biases, and the adjusted estimators are larger than the Naïve estimators.



(a) Separate Data



(b) Combined Data

Figure 5-1 Estimated differences for the Naïve and Adjusted methods

5.2 Observed Count Data

Due to the hotspot identification rule, or warrants for treatments from manuals such as the MUTCD (Department of Transportation et al., 2003), it is not surprising that most crash datasets from current before/after safety studies are truncated by high entry criteria. However, using datasets that only contain a few sample sites make it difficult for a researcher to examine whether our bias-adjusted method works or not. This difficulty stems from the fact that the real population mean, dispersion parameter, and site-selection bias are usually unknown. In other words, we need a better dataset that contains both the *before* and *after* data from the whole population (or as close to it as possible); we can then filter this data and calculate its bias via different entry criteria.

5.2.1 First Dataset

To complete this task, the researcher assumed that there was a dummy treatment applied to 917 sites (which had at least one recorded crash in 2008) in College Station, Texas, on December 31, 2008. This made 2008 the *before* period and 2009 the *after* period. Because there was no such treatment in the real world, its safety effectiveness should be close to 1. The actual safety effectiveness of this dummy treatment (based on the entirety of the crash data available, without selection) was 0.95. We then filtered the crash data by different entry criteria ranging from 1 to 5. Table5-2 shows the Naïve and Adjusted estimators of safety effectiveness, difference, and dispersion parameters in different entry criteria. It was very clear that higher entry criteria results in an overestimation of a treatment's safety effectiveness, especially when using the Naïve method. Also, using the Adjusted method could reduce partial site-selection bias. This also results in the estimators being closer to the true value (0.95). In Figure 5-2(a), the estimators of safety effectiveness (indicated by a green line) are closer to the true value (indicated by a blue line) than when using the Naive method (indicated by a red line), except when the entry criteria are very small. As for other parameters (e.g., the difference and dispersion parameters), the results were analogous to those of safety effectiveness. Setting higher entry criteria causes a higher level of site-selection bias. By removing partial site-selection bias, the Adjusted estimators are usually closer to the true value than the Naïve estimators (Figure 5-2 (b) and Figure 5-2 (c)). Also, it should be noted that the distribution of this dataset followed a negative binomial distribution. For the crash data in the *before* period and *after* period, the inverse dispersion parameters are 1.694 (σ=0.113) and 0.587 (σ=0.037), and the mean parameters are 2.665 (σ=0.086) and 2.530 (σ=0.121) respectively. If the dataset does not follow the assumption distribution (a negative binomial distribution), equations (3.8), (3.11), and (3.16) might not be able to adjust site-selection biases.

Table 5-2 Crash Data for Dummy Treatment in College Station, TX

Safety Effectiveness				Difference			Dispersion Parameter		
entry criteria	θ true	θ naïve	θ adjusted	δ true	δ naïve	δ adjusted	α true	α naïve	α adjusted
>1	0.95	0.96	1.11	-0.13	-0.06	0.16	3.11	2.40	3.30
>2	0.95	0.86	0.94	-0.13	-0.76	-0.12	3.11	1.34	1.85
>3	0.95	0.82	0.89	-0.13	-1.26	-0.42	3.11	1.05	1.59
>4	0.95	0.81	0.85	-0.13	-1.65	-0.62	3.11	0.90	0.83
>5	0.95	0.80	0.85	-0.13	-1.68	-0.65	3.11	0.78	1.18

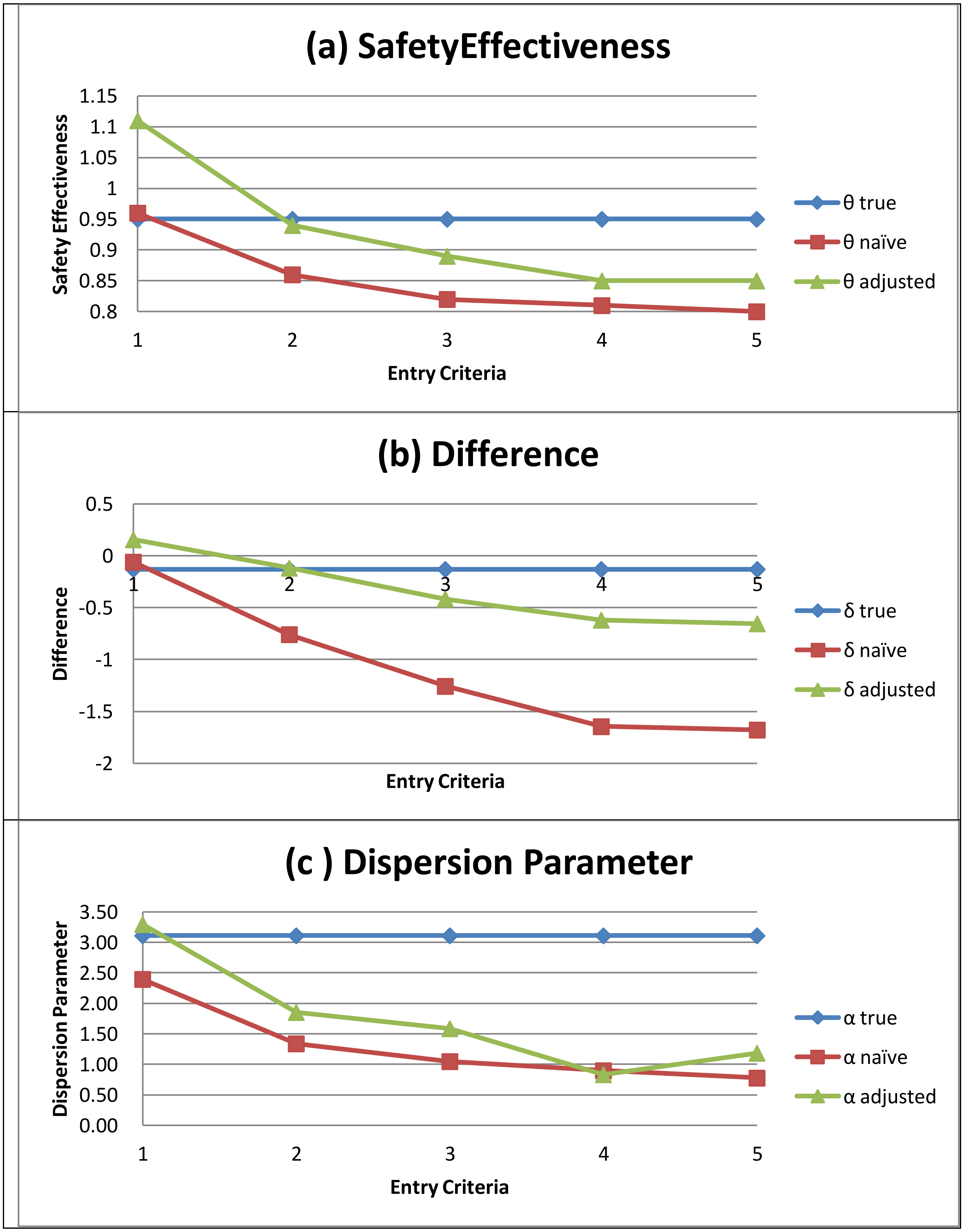


Figure 5-2 Estimated safety effectiveness, difference, and dispersion parameter for the naïve and adjusted methods

5.2.2 Second Dataset

The red-light running-camera crash data were obtained from a Texas Transportation Institute project whose goal is to examine the impact of the installation of such cameras on crash frequency. The original dataset includes 319 intersections in Texas; this dissertation only included 95 intersections, all of which had the same before and after periods (two years) for calculation convenience purposes. This dataset is ideal as an example for examining the Adjusted method, because it offered a large sample size. Additionally, some sites exhibited particularly low crash frequencies (even zero). In other words, its estimators of mean values, dispersion parameters, and indexes of safety effectiveness could be treated as true values because there were no entry criteria for this dataset (the crash frequency of some sites in the *before* period was zero, with the assumption that the dispersion parameter represents the true value). Also, the distribution of the various samples followed our assumption (in that it was a negative binomial). This study assumed that the true values of θ, δ, α were 0.72, -1.93, and 0.77, which was determined by using equations (3.5), (3.9), and (3.10), respectively. The results obtained using the Naïve before-after method showed that overall the crash frequency of all the intersections studied decreased by 6.8%. To examine the effects of the site-selection bias, equation (3.8) was used to obtain the adjusted safety effectiveness. The first entry criterion was assumed to be 1, because the suggested initial assumption for the entry criteria should be considered equal to the smallest set of observed data (i.e., C=min N_{ij}) (Johnson et al., 1970). Then, the minimum entry criteria equal to 1, 2, 3, etc., up to 20 were subsequently employed. The details of the site-selection bias are listed in Table5-3.

Figure 5-3(a) shows the difference between the $\theta_{adjusted}$, θ_{naive}, and the θ_{true}. This figure clearly illustrates that the Adjusted estimators yielded a lower selection bias than the Naïve estimators. Moreover, Figure 5-3(a) shows that higher entry criteria tend to cause higher selection biases, which lead to an overestimation of safety effectiveness. Equation (3.8) partially eliminates selection bias, but does not remove all selection bias (as discussed above). Generally, Figure 5-3(a) supports the theoretical points and simulation results described above; higher criteria cause larger (more negative) biases, and the Adjusted estimators are closer to the true value than that produced by the Naïve estimator. In addition, Table5-3 also shows that higher entry criteria may lead to underestimations of the dispersion parameter and overestimations of the difference. This result is consistent with equations (3.8), (3.11) and (3.15).

Table 5-3 Estimators of Parameters for Observed Crash Data in Different Entry Criteria

Entry Criteria	1	5	10	15	20
θ TRUE	0.72	0.72	0.72	0.72	0.72
θ Naïve	0.72	0.65	0.63	0.62	0.58
θ Adjusted	0.73	0.67	0.65	0.65	0.59
δ TRUE	-1.93	-1.93	-1.93	-1.93	-1.93
δ Naïve	-1.93	-3.68	-5.94	-7.75	-13.50
δ Adjusted	-1.66	-2.50	-4.43	-5.27	-9.43
α True	0.77	0.77	0.77	0.77	0.77
α Naïve	0.77	0.48	0.22	0.21	0.37
α Adjusted	0.80	0.61	0.29	0.30	0.70

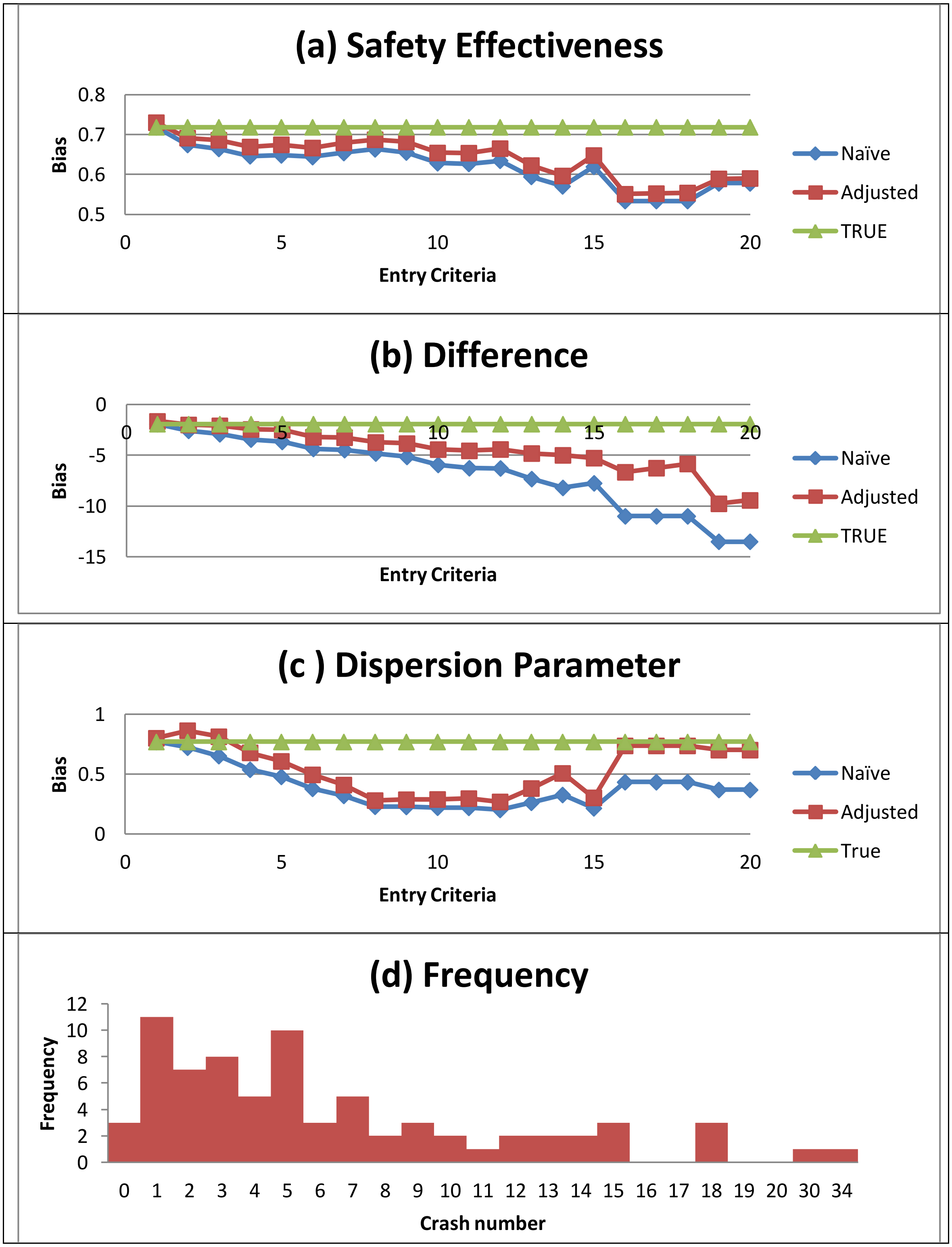


Figure 5-3 Safety effectiveness, difference, and dispersion parameter for the true value, naïve, and adjusted method

5.3 Summary

In this section, the researcher has applied site-section bias estimators to actual continuous data (speed data) and observed count data (crash count data). The results successfully support the researcher’s previous findings based on simulation data: setting higher entry criteria results in higher site-selection bias. Also, the Adjusted estimators were closer to the true value than the Naïve estimators, even for the dispersion parameter. Park and Lord's (2010) study results were used to compare the differences between the adjusted estimators by removing RTM and removing site-selection biases. The Adjusted method can remove all site-selection biases when the dispersion parameter and mean are known. The next section provides a summary and discussion of the research accomplished in this study, and discusses avenues for further work.

6. SUMMARY AND CONCLUSIONS

The before-after study is the most popular method used by traffic engineers and transportation safety analysts for evaluating the effects of an intervention. However, although this kind of study may offer superior performance than cross-sectional study, it can still be plagued by important methodological limitations, which could significantly alter the study outcome. They include the regression-to-the-mean (RTM) and site-selection effects. So far, most of the research on these biases has been directed at the RTM. Hence, the primary objective of this study was to describe how site-selection effects influence the evaluation of treatments. More specifically, the goal was to quantify site-selection bias as a function of different entry criteria and other factors associated with traffic safety data. Then, an Adjusted method was developed to reduce the selection bias when an entry criterion is used in before-after studies without relying on the use of a control group. Moreover, this research also includes the site-selection effects on continuous data (e.g. speed, reaction times, etc.), since the majority of other studies focus on discrete counts. The advantage of using the Adjusted method is that it can adjust the naïve estimator by partially eliminating site-selection bias, even when biased estimators of the mean and dispersion parameter are used. The study objective was accomplished using simulated data (supported by theoretical derivations documented in Appendix A and B) and observed data. Truncated Normal distributions and Truncated Negative Binomial distributions were used to generate the simulated continuous data and count data.

This section highlights the main findings from this research and makes some recommendations for applying the Adjusted method in highway safety research. This dissertation ends with potential topics in which the research can be extended.

6.1 Main Findings

This study has examined how setting an entry criterion influences the estimation of interventions for continuous and count data. The proposed method documented in this research provides a way to adjust the Naïve estimator by using the sample data and without relying on the data collected for the control group, since finding enough appropriate sites for the control group is a common inhibitor to accurate traffic-safety analyses.

In this exercise, the proposed method, a.k.a. the Adjusted method, was compared to commonly used methods in before-after studies. For continuous data, the study results showed that among all methods evaluated (Naïve, CG, ANCOVA, and Adjusted), the Naïve is the most significantly affected by the selection bias. Using a control group (CG) or the ANCOVA method can eliminate site-selection bias, as long as the characteristics of the control group are exactly the same as those for the treatment group. However, control group data that have the same characteristics based on a truncated distribution or sample may not be available in practice. Moreover, site-selection biases generated by using a dissimilar control group might even be higher than wihen using the Naïve method. The Adjusted method, as illustrated in equation (3.4), can partially eliminate site-selection bias even when biased estimators of the mean, variance, and correlation coefficient of a truncated normal distribution are used or are not known with certainty. Based on the simulated scenarios, the study results showed that high entry criteria, high within-subject variance, and small between-subject variance all cause a high site-selection bias. The analysis performed using observed speed data collected in Florida supported the simulation results.

For count data, this study examined estimates of site-selection biases within three parameters: safety effectiveness, difference, and dispersion parameter. For safety effectiveness (θ), the most popular index of countermeasure effectiveness, we obtained results similar to those obtained from the continuous data. Among all the methods evaluated (Naïve, CG, EB_{MM}, and EB_{CG}), the Naïve and EB_{MM} methods are most significantly affected by selection bias. Using the CG or EB_{CG} methods could eliminate site-selection biases if the control group has the exact same mean and dispersion parameter as those for the treatment group or population. In sum, the study revealed that even when the RTM is accounted for (i.e., when using the EB_{MM} method), the index for the safety effectiveness can still be biased when an entry criterion is used, whether it is explicitly defined or not. A theoretical derivation is presented in Appendix A to support the results documented in this study. Based on the simulated scenarios (and also supported theoretically), this study's results showed that higher entry criteria, larger values of the safety index (θ), and smaller dispersion parameter values (α) will all cause a higher site- selection-biased estimate.

However, site-selection bias on safety effectiveness (θ) and difference (δ) are different in two ways. First, using the CG or EB_{CG} method cannot eliminate site-selection bias on the difference (δ) even with a perfect control group, and their estimators of difference are very close to the Naïve and EB_{MM} estimators. A theoretical derivation has been provided to support the above findings. Secondly, according to the simulation result, higher entry criteria, higher dispersion parameters, and higher difference values cause higher site-selection biases on estimator of difference.

Finally, regarding the dispersion parameter, setting higher entry criteria and a higher dispersion parameter might cause higher levels of site-selection bias, which will result in the underestimations of dispersion parameters.

As in the discussion above, the Adjusted method, as illustrated in equations (3.8) and (3.15), can partially eliminate site-selection bias in the estimates of safety effectiveness and difference, even when biased estimators of the mean and dispersion parameter of a truncated Negative Binomial distribution are used or when the crash rate is not a fixed number. However, the Adjusted method does not work well to reduce the site-selection biases in the estimates of dispersion parameters, further study is necessary to solve this problem. The analysis performed using observed crash data collected in Texas supports the simulation results described here.

In sum, this study developed a new adjusted method by combining the study results obtained from Cook and Wei (2002)and Geyer (2007). Specific problems, such as different calculation of the CG method and *dissimilar* control groups, were discussed before we applied these results from medical studies to traffic safety studies. Moreover, two appendices are included here to describe the conditions under which the EB and CG estimators of the index of safety effectiveness (θ) and difference (δ) are asymptotically unbiased and biased, respectively. Also, our results provide evidence to challenge the common assumption that using the EB or the CG method can remove the selection bias. Finally, we hope these study results will help engineers and transportation safety specialists evaluate different treatment countermeasures more efficiently.

6.2 Recommendations and Future Research Areas

Given the nature of the work documented in this dissertation, there are many avenues for further work. First, since this research used a biased estimator for the mean, variance, correlation coefficient, and dispersion parameter to adjust the Naïve estimator, it may be beneficial to apply more advanced techniques to estimate the parameters of a truncated Normal distribution and a truncated Negative Binomial model in an effort to produce more precise estimates. Second, more work needs to be done involving multiple entry criteria in before-after studies, especially when such studies come from different types of data. For instance, a traffic countermeasure might be based on two warrants: a site having more than five crashes in the past year, and where the observed 85% driving speed is over 60 MPH. Third, guidelines should be developed to define what the entry criterion should be when it is not known (e.g., a minimum value, a speed limit, etc.). Finally, a simpler approach for displaying the site-selection biases should be found. Tables based on the sample mean, entry criteria, and level of variance could perhaps be provided in design manuals or similar types of documents. For example, Table 6-1 shows the possible site-selection effects when study sites are chosen according to the top 5% of crash rates. The true values of safety effectiveness and difference are assumed to be 0.5 and -1.5, and the crash rate in the *before* and *after* periods is assumed to be 3 and

1.5, respectively. Traffic engineers can use Table 6-1 to estimate the possible site-selection biases according to the entry criteria and dispersion parameter of their own data.

Table 6-1 Possible Site-Selection Biases in Different Entry Criteria

Entry Criteria	Dispersion Parameter α	μ_1	μ_2	θ_{biased}	δ_{biased}
6	0.3	8.39	2.78	0.33	-5.61
7	0.5	9.56	3.86	0.36	-6.31
8	0.9	11.77	4.7	0.40	-7.07
9	1.2	13.44	5.58	0.42	-7.85
10	1.6	15.3	6.59	0.43	-8.09

In addition, since site-selection bias is a relatively new topic for count data, some issues should be examined in future safety studies. First, one should examine site-selection effects close to the boundary when $\alpha = 0$ as a function of different mean values for the *before* period. Second, statistical tests or methodologies should be developed to ensure that the data collected as the control group for the EB_{CG} method are the same as the full data from which the truncated distribution is used (which may or may not be possible to verify). Although the EB_{CG} method has been (and still is) frequently used among transportation safety analysts, very few ever compare the characteristics of the treatment and control groups. Researchers automatically assume that the NB regression models estimated from the control group have the same characteristics as the sites selected for potential treatment. Also, future studies might apply our Adjusted method to other surrogate measures, such as aggressive lane merging, sharp barking, speeding, and red-light running. Recently these measurements have also been used to estimate the countermeasure effectiveness.

REFERENCES

Abbess, C., 1981. Accidents at blackspots; estimating the effectiveness of remedial treatment with special reference to the regression-to-mean effect, Traffic Engineering & Control 22 (10),535-542.

Barnett, A.G., Van Der Pols, J.C., Dobson, A.J., 2005. Regression to the mean: what it is and how to deal with it, International Journal of Epidemiology, 34 (1), 215-220.

Barr, D.R., Sherrill, E.T., 1999. Mean and variance of truncated normal distributions, American Statistician 53 (4), 357-361.

Chuang-Stein, C., Tong, D.M., 1997. The impact and implication of regression to the mean on the design and analysis of medical investigations, Statistical Methods in Medical Research 6 (2), 115-128.

Cohen Jr, A., 1950. Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples, The Annals of Mathematical Statistics 21(4), 557-569.

Cook, R.J., Wei, W., 2002. Selection effects in randomized trials with count data, Statistics in Medicine 21 (4), 515-531.

Cruyff, M.J.L.F., van der Heijden, P.G.M., 2008. Point and interval estimation of the population size using a zero-truncated negative binomial regression model, Biometrical Journal 50 (6), 1035-1050.

Danielsson, S., 1986. A comparison of two methods for estimating the effect of a countermeasure in the presence of regression effects, Accident Analysis & Prevention 18 (1), 13-23.

Davis, G.A., 2000. Accident reduction factors and causal inference in traffic safety studies: A review, Accident Analysis & Prevention 32 (1), 95-109.

Davis, C.E., 1976. The effect of regression to the mean in epidemiologic and clinical studies, American Journal of Epidemiology 104 (5), 493-498.

Department of Transportation, Federal Highway Administration, 2003. Manual on Uniform Traffic Control Devices. Washington D.C.

Ederer, F., 1972. Serum cholesterol changes: effects of diet and regression toward the mean. Journal of Chronic Disease 25 (5), 277.

Formann, A.K., 2008. Estimating the proportion of studies missing for meta-analysis due to publication bias, Contemporary Clinical Trials 29 (5), 732-739.

Geurts, K., Wets, G., 2003. Black spot analysis methods: literature review, Flemish Research Center for Traffic Safety, Diepenbeek, Belgium .

Geyer, C.J., 2007. Lower-truncated Poisson and negative binomial distributions. University of Minnesota, MN. Accessed Feb. 28th, 2011: http://cran.r-project.org/web/packages/aster/vignettes/trunc.pdf

Harwood, D.W., Bauer, K.M., Richard, K.R., Gilmore, D.K., Graham, J.L., Potts, I.B., Torbic, D.J., Hauer, E., 2007. Methodology to Predict the Safety Performance of Urban and Suburban Arterials. NCHRP Web Document No. 129: Phases I and II, Transportation Research Board, Washington, DC

Hauer, E., 1997. Observational Before--After Studies in Road Safety, Elsevier Science Ltd., Oxford.

Hauer, E., 1980a. Bias-by-selection: overestimation of the effectiveness of safety countermeasures caused by the process of selection for treatment, Accident Analysis & Prevention 12 (2), 113-117.

Hauer, E., 1980b. Selection for treatment as a source of bias in before-and-after studies, Traffic Engineering and Control 21 (8), 419-421.

Hauer, E., Byer, P., Joksch, H., 1983. Bias-by-selection: the accuracy of an unbiased estimator, Accident Analysis & Prevention 15 (5), 323-328.

Hauer, E., Persaud, B., 1987. How to estimate the safety of rail-highway grade crossings and the safety effects of warning devices. Transportation Research Record 1114, 31–140

Hauer, E., 1984. Problem of identifying hazardous locations using accident data. Transportation Research Record 975, 36–43

Hauer, E., Byer, P., Joksch, H.C., 1983. Bias-by-selection: The accuracy of an unbiased estimator, Accident Analysis & Prevention 15 (5), 323-328.

Heydecker, B., Wu, J., 2001. Identification of sites for road accident remedial work by Bayesian statistical methods: an example of uncertain inference, Advances in Engineering Software 32 (10-11), 859-869.

Higle, J.L., Witkowski, J.M., 1988. Bayesian identification of hazardous locations., Transportation Research Record 1185, 24–36.

Johnson, N.L., Kotz, S., Balakrishnan, N., 1970. Distributions in statistics: continuous univariate distributions-1, Boston: Houghton Miffin Company.

Kennedy, B.S., 2005. Does race predict stroke readmission? An analysis using the truncated negative binomial model. Journal of the National Medical Association 97 (5), 699.

Lee, A.H., Wang, K., Yau, K.K.W., Somerford, P.J., 2003. Truncated negative binomial mixed regression modelling of ischaemic stroke hospitalizations, Statistics in Medicine 22 (7), 1129-1139.

Li, W., Carriquiry, A., Pawlovich, M., Welch, T., 2008. The choice of statistical models in road safety countermeasure effectiveness studies in Iowa, Accident Analysis & Prevention 40 (4), 1531-1542.

Lord, D., 2006. Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter, Accident Analysis & Prevention 38 (4), 751-766.

Lord, D., and P-F. Kuo. Examining the Effects of Site Selection Criteria for Evaluating the Effectiveness of Traffic Safety Improvement Countermeasures. 90th Annual Meeting of the Transportation Research Board, Washington, D.C., 2009.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives, Transportation Research Part A: Policy and Practice 44 (5), 291-305.

Maher, M., Mountain, L., 2009. The sensitivity of estimates of regression to the mean, Accident Analysis & Prevention 41 (4), 861-868..

Miaou, S.P., Song, J.J., 2005. Bayesian ranking of sites for engineering safety improvements: decision parameter, treatability concept, statistical criterion, and spatial dependence, Accident Analysis & Prevention 37 (4), 699-720.

Miranda-Moreno, L.F., 2006. Statistical models and methods for identifying hazardous locations for safety improvements, Ph.D. Dissertation. Department of Civil Engineering, University of Waterloo, Waterloo, Ontario, Canada..

Miranda-Moreno, L.F., Fu, L., Ukkusuri, S., Lord, D., 2009. How to incorporate accident severity and vehicle occupancy into the hot spot identification process? Transportation Research Record 2102 (1), 53-60.

Muchuruza, V., Mussa, R.N., 2004. Speeds on rural interstate highways relative to posting the 40 mph minimum speed limit, Journal of Transportation and Statistics 7 (2/3).

Noland, R.B., 2003. Traffic fatalities and injuries: the effect of changes in infrastructure and other trends, Accident Analysis & Prevention 35 (4), 599-611.

Park, E.S., Park, J., Lomax, T.J., 2010. A fully Bayesian multivariate approach to before-after safety evaluation, Accident Analysis & Prevention 42 (4), 1118-1127.

Park, P.Y., Lord, D., 2010. Investigating regression to the mean in before-and-after speed data analysis, Transportation Research Record: Journal of the Transportation Research Board 2165 (1), 52-58.

Pendleton, O.J., 1991. Application of new accident analysis methodologies volume i–general methodology, Office of research and development, Turner-Fairband Highway Research Centre, U.S. Department of Transportation. .

Persaud, B., Lyon, C., 2007. Empirical Bayes before-after safety studies: Lessons learned from two decades of experience and future directions, Accident Analysis & Prevention 39 (3), 546-555.

Persaud, B., Lyon, C., Nguyen, T., 1999. Empirical Bayes procedure for ranking sites for safety investigation by potential for safety improvement, Transportation Research Record: Journal of the Transportation Research Board 1665 (1), 7-12.

Persaud, B.N., 2001. Statistical Methods in Highway Safety Analysis. Transportation Research Board, National Research Council, Washington, D.C..

Robbins, H., 1964. An empirical Bayes approach to statistics, The Annals of Mathematical Statistics 35 (1), 1-20

Schlüter, P., Deely, J., Nicholson, A., 1997. Ranking and selecting motor vehicle accident sites by using a hierarchical Bayesian model, Journal of the Royal Statistical Society: Series D (The Statistician) 46 (3), 293-316.

Stigler, S.M., 1997. Regression towards the mean, historically considered, Statistical Methods in Medicine Research 6 (2), 103.

Tarko, A., Eranky, S., Sinha, K., 1998. Methodological considerations in the development and use of crash reduction factors. Paper presented at 77th meeting of Transportation Research Board, Washington, DC

Wright, C., Abbess, C., Jarrett, D., 1988. Estimating the regression-to-mean effect associated with road accident black spot treatment: towards a more realistic approach, Accident Analysis & Prevention 20 (3), 199-214.

Ye, Z., Lord, D., 2009. Estimating the variance in before-after studies, Journal of Safety Research 40 (4), 257-263.

Yudkin, P., Stratton, I., 1996. How to deal with regression to the mean in intervention studies, Lancet 347 (8996), 241-243.

APPENDIX A

This appendix describes the conditions under which the empirical Bayes (EB) estimator of the index of safety effectiveness (θ) is asymptotically unbiased and biased. The bias is defined as the difference between $\theta = \frac{\Lambda_2}{\Lambda_1}$ and the expected value of $\hat{\theta}$. The following paragraphs show the EB estimators for three different cases: (1) Without entry criteria; (2) With entry criteria and with "perfect"[2] control group data; and, (3) With entry criteria but without "perfect"[2] control group data. The first one is the most common estimator, and previous studies have already shown that it is unbiased (Robbins, 1956; Hauer, 1997). For the second estimator, it is also unbiased, and the results are consistent with the Davis (2000) study. For the third estimator, (EB_{MM}), which is the one used in this research, we demonstrate when the estimator can be as biased in the same way as when using the Naïve method. To simplify the comparison, all three estimators are shown below.

It should be noted that the moment estimators, maximum likelihood estimators, or other estimators based on conditional data consistently estimate $E(\hat{\theta}) = \frac{E(N_{i2} | N_{i1} > C)}{E(N_{i1} | N_{i1} > C)}$ rather than $\hat{\theta} = \frac{E(N_{i2})}{E(N_{i2})} = \frac{\Lambda_2}{\Lambda_1}$. All notations in this appendix are the same as in the main text, and $\mu_1 = E\left(N_{i1} | N_{i1} > C\right)$ is for notational convenience.

[2]: The perfect reference or control group has the exact same mean and dispersion parameter (or the variance) as those of the treatment group. This situation rarely happens occurs in the real world, because the true mean of the control group is most likely unknown.

1. <u>Experiments without effective entry criteria</u>

For this case, the crash frequency of site i in the before period (N_{i1}) can be any non-negative integers (e.g. 0, 1, 2, …). When there are no entry criteria, C is equal to -1 or less.

The EB estimator for Case 1 is given as follows:

$$\theta = \frac{\Lambda_2}{\Lambda_1}$$

$$\hat{\theta}_{EB} = \frac{\Lambda_2}{\Lambda_{1(EB)}} = \frac{\frac{1}{m}\sum_{i=1}^{m}\left(N_{i2} \middle| N_{i1}\right)}{\frac{1}{m}\sum_{i=1}^{m}\left[\left(\frac{1}{1+\Lambda_1\alpha}\right)\Lambda_1 + \left(\frac{\Lambda_1\alpha}{1+\Lambda_1\alpha}\right)N_{i1}\right]} \qquad (\Lambda_1 \textit{ and } \alpha \textit{ are unknown})$$

$$= \frac{\frac{1}{m}\sum_{i=1}^{m}\left(N_{i2} \middle| N_{i1}\right)}{\left(\frac{1}{1+\Lambda_1\alpha}\right)\Lambda_1 + \left(\frac{\Lambda_1\alpha}{1+\Lambda_1\alpha}\right)\frac{1}{m}\sum_{i=1}^{m}N_{i1}}$$

Cook and Wei (2002) show that $\frac{1}{m}\sum_{i=1}^{m}\left(N_{i2} \mid N_{i1}\right)$ converges in probability to $\Lambda_2\left(\frac{N_1\alpha+1}{\Lambda_1\alpha+1}\right)$ and $\frac{1}{m}\sum_{i=1}^{m}N_{i1}$ converges in probability to Λ_1.

Likewise, $\hat{\alpha}$ converges in probability to α, $\hat{\Lambda}_1$ converges in probability to Λ_1

Therefore, $\widehat{\theta_{EB}}$ converges in probability to

$$= \frac{\Lambda_2\left(\frac{N_1\alpha+1}{\Lambda_1\alpha+1}\right)}{\left(\frac{1}{1+\Lambda_1\alpha}\right)\Lambda_1 + \left(\frac{\Lambda_1\alpha}{1+\Lambda_1\alpha}\right)\Lambda_1}$$

$$= \frac{\Lambda_2}{\Lambda_1} = \theta$$

(A-1)

Because there are no entry criteria, the expected value of the crash count in the *before* period is equal to the long-term mean crash rate for all the sites in the sample population in the *before* period. ($E(N_{i1}) = \mu_1 = \Lambda_1$) Also, the expected value of the crash count in the *after* period is equal to $\Lambda_2\left(\frac{\mu_1\alpha+1}{\Lambda_1\alpha+1}\right)$. Hence, the EB estimator without entry criteria is unbiased. Please see the appendix of the paper written by Cook and Wei (2002) for additional details about this proof.

2. <u>Experiments with entry criteria and a perfect control group</u>

For Case 2, the number of crashes can be any integers larger than 0, such that $N_{i1}>C=0$ (N_{i1} = 1, 2, 3, …). The perfect reference group data are used to estimate the dispersion parameter and the mean, which have the same values as those for the treatment group: α, Λ_1.

The EB estimator for Case 2 is given as follows:

$$\theta = \frac{\Lambda_2}{\Lambda_1}$$

$$\hat{\theta}_{EB} = \frac{\Lambda_2}{\Lambda_{1(EB)}} = \frac{\frac{1}{m}\sum_{i=1}^{m}\left(N_{i2} \middle| N_{i1} > C\right)}{\frac{1}{m}\sum_{i=1}^{m}\left[\left(\frac{1}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\hat{\Lambda}_1 + \left(\frac{\hat{\Lambda}_1\hat{\alpha}}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\left(N_{i1} \middle| N_{i1} > C\right)\right]} \qquad (\Lambda_1 \textit{ and } \alpha \textit{ are unknown})$$

$$= \frac{\frac{1}{m}\sum_{i=1}^{m}\left(N_{i2} \middle| N_{i1} > C\right)}{\left(\frac{1}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\hat{\Lambda}_1 + \left(\frac{\hat{\Lambda}_1\hat{\alpha}}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\frac{1}{m}\sum_{i=1}^{m}(N_{i1} | N_{i1} > C)}$$

Cook and Wei show that $\frac{1}{m}\sum_{i=1}^{m}\left(N_{i2} \mid N_{i1} > C\right)$ converges in probability

to $\Lambda_2\left(\frac{\mu_1\alpha+1}{\Lambda_1\alpha+1}\right)$.

For the EB control group method, $\hat{\alpha}$ converges in probability to α, $\hat{\Lambda}_1$ converges in probability to Λ_1. Therefore, $\widehat{\theta_{EB\_CG}}$ converges in probability to

$$= \frac{\Lambda_2\left(\frac{\mu_1\alpha+1}{\Lambda_1\alpha+1}\right)}{\left(\frac{1}{1+\Lambda_1\alpha}\right)\Lambda_1 + \left(\frac{\Lambda_1\alpha}{1+\Lambda_1\alpha}\right)\mu_1} \qquad (\Lambda_1 \textit{ may not be equal to } \mu_1)$$

$$= \frac{\Lambda_2\left(\frac{\mu_1\alpha+1}{\Lambda_1\alpha+1}\right)}{\left(\frac{\Lambda_1}{1+\Lambda_1\alpha}\right)(1+\mu_1\alpha)} = \frac{\Lambda_2}{\Lambda_1} = \theta$$

(A-2)

The long-term mean and dispersion parameter values are estimated using a control group or regression model based on the control group. As in the first case, the EB estimator is unbiased. However, in practice, the characteristics of the control group may not be the same as the one used for the treatment group.

3. Experiments with entry criteria and without perfect control group data.

   For this case, the number of crashes can be any integers larger than 0, such that $N_{i1}>C=0$ (N_{i1} = 1, 2, 3, …). The reference group data are used to estimate the dispersion parameter and the mean, which have different values than those from the treatment group.

   The EB estimator for Case 3, based on the method of moment, is given as follows:

$$\hat{\theta}_{EB} = \frac{\Lambda_2}{\Lambda_{1(EB)}} = \frac{\frac{1}{m}\sum_{i=1}^{m}\left(N_{i2} \mid N_{i1} > C\right)}{\frac{1}{m}\sum_{i=1}^{m}\left[\left(\frac{1}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\hat{\Lambda}_1 + \left(\frac{\hat{\Lambda}_1\hat{\alpha}}{1+\hat{\Lambda}_1\hat{\alpha}}\right)(N_{i1} \vert N_{i1} > C)\right]} \qquad (\Lambda_1 \textit{ and } \alpha \textit{ are unknown})$$

$$= \frac{\frac{1}{m}\sum_{i=1}^{m}\left(N_{i2} \vert N_{i1} > C\right)}{\left(\frac{1}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\hat{\Lambda}_1 + \left(\frac{\hat{\Lambda}_1\hat{\alpha}}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\frac{1}{m}\sum_{i=1}^{m}\left(N_{i1} \vert N_{i1} > C\right)}$$

Cook and Wei (2002) show that $\frac{1}{m}\sum_{i=1}^{m}\left(N_{i2} \mid N_{i1} > C\right)$ converges in probability to $\Lambda_2\left(\frac{\mu_1\alpha+1}{\Lambda_1\alpha+1}\right)$.

For the method of moment, we assume $\hat{\Lambda}_1$ is equal to $\frac{1}{m}\sum_{i=1}^{m}\left(N_{i1} \vert N_{i1} > C\right)$ which converges in probability to $E(N_{i1} \vert N_{i1} > C)$.

Therefore, $\widehat{\theta_{EB\_MM}}$ converges in probability to

$$= \frac{\Lambda_2\left(\frac{\mu_1\alpha+1}{\Lambda_1\alpha+1}\right)}{\left(\frac{1}{1+\hat{\Lambda}_1\hat{\alpha}}\right)E(N_{i1} \vert N_{i1} > C) + \left(\frac{\hat{\Lambda}_1\hat{\alpha}}{1+\hat{\Lambda}_1\hat{\alpha}}\right)E(N_{i1} \vert N_{i1} > C)}$$

$$= \frac{\Lambda_2\left(\frac{\mu_1\alpha+1}{\Lambda_1\alpha+1}\right)}{\mu_1}$$

Geyer (2007) shows that μ_1 is equal to $\Lambda_1 + \frac{C+1}{\left(\frac{1}{1+\Lambda_1\alpha}\right)\times\left(1+\frac{P(N_{i1}>C)}{P(N_{i1}=C)}\right)}$.

(A-3)

Equation (A-3) shows that the EB estimator is biased, until there are no entry criteria(C=-1 and $\mu_1 = \Lambda_1$) or $\alpha \to \infty$. Also, the EB estimator for the index of effectiveness (θ) is actually equal to the Naïve before-after estimator, as shown below:

$$\hat{\theta}_{EB\_MM} = \frac{\sum_{i=1}^{m}\left(N_{i2} \mid N_{i1} > C\right)}{\sum_{i=1}^{m}\left[\left(\frac{1}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\hat{\Lambda}_1 + \left(\frac{\hat{\Lambda}_1\hat{\alpha}}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\left(N_{i1} \middle| N_{i1} > C\right)\right]}$$

$$= \frac{\sum_{i=1}^{m}\left(N_{i2} \middle| N_{i1} > C\right)}{\left(\frac{m}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\times\frac{\sum_{i=1}^{m}\left(N_{i1} \middle| N_{i1} > C\right)}{m} + \left(\frac{\hat{\Lambda}_1\hat{\alpha}}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\times\sum_{i=1}^{m}\left(N_{i1} \middle| N_{i1} > C\right)}$$

$$= \frac{\sum_{i=1}^{m}\left(N_{i2} \middle| N_{i1} > C\right)}{\sum_{i=1}^{m}\left(N_{i1} \middle| N_{i1} > C\right)} \tag{A-4}$$

In sum, using a biased estimate of $\hat{\Lambda}_1$ is the main reason why the EB_{mm} estimator of θ is also biased.

APPENDIX B

This appendix describes the conditions under which the control group and empirical Bayes (EB) estimators of the index of safety effectiveness (δ) are asymptotically unbiased and biased. The bias is defined as the difference between $\delta = \Lambda_2 - \Lambda_1$ and the expected value of δ. The following paragraphs show the CG and EB estimators for three different cases: (1) Without entry criteria; (2) With entry criteria and with "perfect"[2] control group data; and, (3) With entry criteria but without "perfect"[3] control group data. The first one is the most common estimator, and previous studies have already shown that it is unbiased (Robbins, 1956; Hauer, 1997). For the second estimator, it is biased. For the third estimator (CG_{un}, EB_{MM}), which is the one used in this research, we demonstrate when the estimator can be as biased as for the naïve method. To simplify the comparison, all three estimators are shown below.

It should be noted that the moment estimators, maximum likelihood estimators, or other estimators based on conditional data consistently estimate $E(\hat{\delta}) = E(N_{i2} | N_{i1} > C) - E(N_{i1} | N_{i1} > C)$ rather than $\hat{\theta} = E(N_{i2}) - E(N_{i2}) = \Lambda_2 - \Lambda_1$. All notations in this appendix are same as those in the main text, and $\mu_1 = E\left(N_{i1} | N_{i1} > C\right)$ for notational convenience.

1. <u>Experiments without effective entry criteria but with a perfect control group</u>

   For this case, the crash frequency of site i in the before period (N_{i1}) can be any non-negative integers (e.g. 0, 1, 2, …). When there are no entry criteria, C is equal to -1 or less.

   The CG and EB estimators for Case 1 are given as follows:

2: The perfect reference or control group has the exact same mean and dispersion parameter (or the variance) as those of the treatment group. This situation rarely happens in the real world, because the true mean of the control group is most likely unknown.

$$\delta = \Lambda_2 - \Lambda_1$$

$$\hat{\delta}_{CG} = \Lambda_2 - \Lambda_{1(CG)} = \Lambda_2^T - \Lambda_1^T \times \frac{\Lambda_2^C}{\Lambda_1^C}$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left(N^T{}_{i2}\middle|N^T{}_{i1}\right) - \frac{1}{m}\sum_{i=1}^{m}\left(N^T{}_{i1}\middle|N^T{}_{i1}\right) \times \frac{\frac{1}{m}\sum_{i=1}^{m}\left(N^C{}_{i2}\middle|N^C{}_{i1}\right)}{\frac{1}{m}\sum_{i=1}^{m}\left(N^C{}_{i1}\middle|N^C{}_{i1}\right)}$$

Cook and Wei (2002) show that $\frac{1}{m}\sum_{i=1}^{m}\left(N_{i2} \mid N_{i1}\right)$ converges in probability to $\Lambda_2\left(\frac{N_1\alpha+1}{\Lambda_1\alpha+1}\right)$ and $\frac{1}{m}\sum_{i=1}^{m}N_{i1}$ converges in probability to Λ_1.

Therefore, $\hat{\delta}_{CG}$ converges in probability to

$$= \Lambda_2\left(\frac{N_1\alpha+1}{\Lambda_1\alpha+1}\right) - \Lambda_1 \times \frac{\Lambda_2\left(\frac{N_1\alpha+1}{\Lambda_1\alpha+1}\right)}{\Lambda_1}$$

$$= \Lambda_2 - \Lambda_1 = \delta$$

$$\hat{\delta}_{EB} = \widehat{\Lambda_2 - \Lambda_{1(EB)}} = \frac{1}{m}\sum_{i=1}^{m}\left(N_{i2}\middle|N_{i1}\right) - \left[\left(\frac{1}{1+\widehat{\Lambda_1\alpha}}\right)\widehat{\Lambda_1} + \left(\frac{\widehat{\Lambda_1\alpha}}{1+\widehat{\Lambda_1\alpha}}\right)\frac{1}{m}\sum_{i=1}^{m}N_{i1}\right]$$

Same as above, $\frac{1}{m}\sum_{i=1}^{m}\left(N_{i2} \mid N_{i1}\right)$ converges in probability to $\Lambda_2\left(\frac{N_1\alpha+1}{\Lambda_1\alpha+1}\right)$ and $\frac{1}{m}\sum_{i=1}^{m}N_{i1}$ converges in probability to Λ_1.

Likewise, $\hat{\alpha}$ converges in probability to α, $\hat{\Lambda}_1$ converges in probability to Λ_1

Therefore, $\hat{\delta}_{EB}$ converges in probability to

$$= \Lambda_2\left(\frac{N_1\alpha+1}{\Lambda_1\alpha+1}\right) - \left[\left(\frac{1}{1+\Lambda_1\alpha}\right)\Lambda_1 + \left(\frac{\Lambda_1\alpha}{1+\Lambda_1\alpha}\right)\Lambda_1\right]$$

$$= \Lambda_2 - \Lambda_1 = \delta$$

(B-1)

Because there are no entry criteria, the expected value of the crash count in the *before* period is equal to the long-term crash mean for all the sites in the sample population in the *before* period. ($E(N_{i1}) = \mu_1 = \Lambda_1$) Also, the expected value of the crash count in the *after* period is equal to $\Lambda_2 \left(\frac{\mu_1 \alpha + 1}{\Lambda_1 \alpha + 1} \right)$. Hence, the CG and EB estimators without entry criteria are unbiased. Please see the appendix of the paper written by Cook and Wei (2002) for additional details about this proof.

2. <u>Experiments with entry criteria and a perfect control group</u>

For Case 2, the number of crashes can be any integers larger than 0, such that N_{i1}>C=0 (N_{i1} = 1, 2, 3, …). The perfect reference group data are used to estimate the dispersion parameter and the mean, which have the same values as those for the treatment group: α, Λ_1.

The CG and EB estimators for Case 2 are given as follows:

$$\delta = \Lambda_2 - \Lambda_1$$

$$\hat{\delta}_{CG} = \Lambda 2 - \Lambda 1(CG) = \Lambda_2^T - \Lambda_1^T \times \frac{\Lambda_2^C}{\Lambda_1^C}$$

$$= \frac{1}{m}\sum_{i=1}^{m}\left(N^T{}_{i2}\left|N_{i1} > C\right.\right) - \frac{1}{m}\sum_{i=1}^{m}\left(N^T{}_{i1}\left|N^T{}_{i1} > C\right.\right) \times \frac{\frac{1}{m}\sum_{i=1}^{m}\left(N^C{}_{i2}\left|N^C{}_{i1} > C\right.\right)}{\frac{1}{m}\sum_{i=1}^{m}\left(N^C{}_{i1}\left|N^C{}_{i1} > C\right.\right)}$$

Cook and Wei show that $\frac{1}{m}\sum_{i=1}^{m}\left(N_{i2} \mid N_{i1} > C\right)$ converges in probability

to $\Lambda_2\left(\frac{\mu_1\alpha + 1}{\Lambda_1\alpha + 1}\right)$.

Therefore, $\hat{\delta}_{CG}$ converges in probability to

$$= \Lambda^T{}_2\left(\frac{\mu^T{}_1\alpha + 1}{\Lambda^T{}_1\alpha + 1}\right) - \mu^T{}_1 \frac{\Lambda^C{}_2\left(\frac{\mu^C{}_1\alpha + 1}{\Lambda^C{}_1\alpha + 1}\right)}{\mu^C{}_1} \qquad (\Lambda^T{}_1 = \Lambda^C{}_1 = \Lambda^C{}_2 = \Lambda_1,\ \mu^T{}_1 = \mu^C{}_1 = \mu_1)$$

$$= \frac{(\Lambda_2 - \Lambda_1)(\mu_1\alpha + 1)}{\Lambda_1\alpha + 1} \neq \Lambda_2 - \Lambda_1 = \theta$$

$$\hat{\delta}_{EB} = \widehat{\Lambda 2 - \Lambda 1(EB)} = \frac{1}{m}\sum_{i=1}^{m}\left(N_{i2}\left|N_{i1} > C\right.\right) - \left[\left(\frac{1}{1 + \hat{\Lambda}_1\hat{\alpha}}\right)\hat{\Lambda}_1 + \left(\frac{\hat{\Lambda}_1\hat{\alpha}}{1 + \hat{\Lambda}_1\hat{\alpha}}\right)\frac{1}{m}\sum_{i=1}^{m}(N_{i1}\left|N_{i1} > C\right.)\right]$$

Same as above, $\frac{1}{m}\sum_{i=1}^{m}\left(N_{i2} \mid N_{i1} > C\right)$ converges in probability to $\Lambda_2\left(\frac{\mu_1\alpha + 1}{\Lambda_1\alpha + 1}\right)$.

For the EB control group method, $\hat{\alpha}$ converges in probability to $\alpha, \hat{\Lambda}_1$ converges in probability to Λ_1.

Therefore, $\hat{\delta}_{EB}$ converges in probability to

$$= \Lambda_2\left(\frac{\mu_1\alpha + 1}{\Lambda_1\alpha + 1}\right) - \left[\left(\frac{1}{1 + \Lambda_1\alpha}\right)\Lambda_1 + \left(\frac{\Lambda_1\alpha}{1 + \Lambda_1\alpha}\right)\mu_1\right] \qquad (\Lambda_1 \textit{ may not be equal to } \mu_1)$$

$$= (\Lambda_2 - \Lambda_1)\left(\frac{\mu_1\alpha + 1}{\Lambda_1\alpha + 1}\right) \neq \Lambda_2 - \Lambda_1 = \delta$$

(B-2)

The long-term mean and dispersion-parameter values are estimated using a control group or regression model based on the control group. Unlike the first case, the CG and EB estimators of difference are biased. Also, it should be noted that the CG and EB estimators of safety effectiveness are unbiased.

3. Experiments with entry criteria and without perfect control group data

For this case, the number of crashes can be any integers larger than 0, such that $N_{i1}>C=0$ (N_{i1} = 1, 2, 3, …). The reference group data are used to estimate the dispersion parameter and the mean, which have the different values than those from the treatment group.

The CG and EB estimators for Case 3, based on the method of moment, are given as follows:

$\delta = \Lambda_2 - \Lambda_1$

$\hat{\delta}_{CG} = \Lambda_2 - \Lambda_{1(CG)} = \Lambda_2^T - \Lambda_1^T \times \dfrac{\Lambda_2^C}{\Lambda_1^C}$

Same as above, $\hat{\delta}_{CG}$ converges in probability to

$$= \Lambda^T{}_2\left(\frac{\mu^T{}_1\alpha+1}{\Lambda^T{}_1\alpha+1}\right) - \mu^T{}_1 \frac{\Lambda^C{}_2\left(\dfrac{\mu^C{}_1\alpha+1}{\Lambda^C{}_1\alpha+1}\right)}{\mu^C{}_1} \neq (\Lambda_2 - \Lambda_1 = \delta)$$

Because the control group is not perfect, so $\Lambda^T{}_1 \neq \Lambda^C{}_1 = \Lambda^C{}_2$, $\mu^T{}_1 \neq \mu^C{}_1$.

$$\hat{\delta}_{EB} = \widehat{\Lambda_2 - \Lambda_{1(EB)}} = \frac{1}{m}\sum_{i=1}^{m}\left(N_{i2} \middle| N_{i1} > C\right) - \left[\left(\frac{1}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\hat{\Lambda}_1 + \left(\frac{\hat{\Lambda}_1\hat{\alpha}}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\frac{1}{m}\sum_{i=1}^{m}\left(N_{i1} \middle| N_{i1} > C\right)\right]$$

Cook and Wei (2002) show that $\frac{1}{m}\sum_{i=1}^{m}\left(N_{i2} \mid N_{i1} > C\right)$ converges in probability

to $\Lambda_2\left(\dfrac{\mu_1\alpha+1}{\Lambda_1\alpha+1}\right)$.

For the method of moment, we assume $\hat{\Lambda}_1$ is equal to $\frac{1}{m}\sum_{i=1}^{m}\left(N_{i1} \middle| N_{i1} > C\right)$

which converges in probability to $E(N_{i1}|N_{i1} > C)$.

Therefore, $\widehat{\theta_{EB\_MM}}$ converges in probability to

$$= \Lambda_2\left(\frac{\mu_1\alpha+1}{\Lambda_1\alpha+1}\right) - \left[\left(\frac{1}{1+\hat{\Lambda}_1\alpha}\right)E(N_{i1}|N_{i1} > C) + \left(\frac{\hat{\Lambda}_1\hat{\alpha}}{1+\hat{\Lambda}_1\alpha}\right)E(N_{i1}|N_{i1} > C)\right]$$

$$= \Lambda_2\left(\frac{\mu_1\alpha+1}{\Lambda_1\alpha+1}\right) - \mu_1 \neq (\Lambda_2 - \Lambda_1 = \delta)$$

Geyer (2007) shows that μ_1 is equal to $\Lambda_1 + \dfrac{C+1}{\left(\dfrac{1}{1+\Lambda_1\alpha}\right)\times\left(1+\dfrac{P(N_{i1} > C)}{P(N_{i1} = C)}\right)}$.

(B-3)

Equation (B-3) shows that the CG and EB estimators are biased until there are no entry criteria (C=-1 and $\mu_1 = \Lambda_1$). It should be noted that the selection bias of difference still exists when α is close to ∞. Also, the EB estimator for the index of effectiveness (δ) is actually equal to the Naïve before-after estimator, as shown below:

$$\hat{\delta}_{EB\_MM}=\sum_{i=1}^{m}\left(N_{i2}\mid N_{i1}>C\right)-\sum_{i=1}^{m}\left[\left(\frac{1}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\hat{\Lambda}_1+\left(\frac{\hat{\Lambda}_1\hat{\alpha}}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\left(N_{i1}\middle|N_{i1}>C\right)\right]$$

$$=\sum_{i=1}^{m}\left(N_{i2}\middle|N_{i1}>C\right)-\left[\left(\frac{m}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\times\frac{\sum_{i=1}^{m}\left(N_{i1}\middle|N_{i1}>C\right)}{m}+\left(\frac{\hat{\Lambda}_1\hat{\alpha}}{1+\hat{\Lambda}_1\hat{\alpha}}\right)\times\sum_{i=1}^{m}\left(N_{i1}\middle|N_{i1}>C\right)\right]$$

$$=\sum_{i=1}^{m}\left(N_{i2}\middle|N_{i1}>C\right)-\sum_{i=1}^{m}\left(N_{i1}\middle|N_{i1}>C\right)=\hat{\delta}_{Naive}$$

(B-4)

In sum, using a biased estimate of $\hat{\Lambda}_1$ is the main reason why the EB_{mm} estimator of θ is also biased.

VITA

Name: Pei-Fen Kuo

Address: Zachry Department of Civil Engineering
Texas A&M University
3136 TAMU
College Station, Texas 77843-3136

Email Address: peifenkuo@gmail.com

Education: B.A., Civil Engineering, National Taiwan University, 2001
M.S., Civil Engineering, National Taiwan University, 2003
Ph.D, Civil Engineering, Texas A&M University, 2012